

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 1, 2005, 08:15:30 ; Search time 189 Seconds
(without alignments)
378.935 Million cell updates/sec

Title: US-09-924-340-58

Sequence: 1 MGPPEFGKGTGHPGLGPKG.....GAMPMEQGYPMKTMKGPFG 163

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database :
1: A.GeneSeq_21:*
2: geneSeqp1980s:*
3: geneSeqp1990s:*
4: geneSeqp2000s:*
5: geneSeqp2001s:*
6: geneSeqp2002s:*
7: geneSeqp2003as:*
8: geneSeqp2004s:*
9: geneSeqp2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	952	100.0	163	6	ABR48481
2	840.5	88.3	1603	8	ADQ21405
3	620	65.1	218	7	ADL25758
4	419	44.0	330	2	AAW57645
5	419	44.0	408	2	AAW07539
6	419	44.0	408	4	ABR64007
7	419	44.0	408	4	AAW2737
8	412.5	43.3	357	2	AAW5115
9	412.5	43.3	357	2	AAW57646
10	412.5	43.3	357	4	AAW64008
11	412.5	43.3	357	4	AAW72738
12	399.5	42.0	1071	7	ADP94310
13	395	41.5	67	5	ABR01788
14	393.5	41.3	503	9	ADW58908
15	390.5	41.0	553	8	ADP69307
16	390	41.0	546	2	AAW40115
17	390	41.0	632	5	ABP43633
18	390	41.0	910	8	ADQ19402
19	390	41.0	1268	8	ABO84442
20	390	41.0	1283	8	ABO84443
21	390	41.0	1685	4	ABG044839
22	390	41.0	1685	8	ABO84444
23	390	41.0	1693	4	ABG15619
24	389.5	40.9	1064	2	AAW93254

25	389.5	40.9	1064	2	AAW57652	AAW57652 Collagen-
26	389.5	40.9	1065	2	AAW37741	AAW37741 Collagen-
27	389.5	40.8	561	2	AAW37739	AAW37739 Collagen-
28	388.5	40.8	561	2	AAW93249	AAW93249 Collagen-
29	388.5	40.8	561	2	AAW57650	AAW57650 Collagen-
30	388	40.8	1497	5	ABP43711	ABP43711 Bullous p
31	386.5	40.6	842	8	ABO84441	ABO84441 Mouse can
32	386	40.5	252	2	AAW37743	AAW37743 Collagen-
33	385	40.4	252	2	AAW37738	AAW37738 Collagen-
34	385	40.4	252	2	AAW93248	AAW93248 Collagen-
35	385	40.4	1497	6	ADW83846	ADW83846 Human COL
36	384	40.3	1466	4	AAE02534	AAE02534 Bovine al
37	384	40.3	1466	4	AAE02533	AAE02533 Bovine al
38	383.5	40.3	252	2	AAW01418	AAW01418 Collagen-
39	383.5	40.3	1838	2	AAW53257	AAW53257 Human COL
40	383.5	40.3	1838	7	ADW55566	ADW55566 Human COL
41	383.5	40.3	1838	7	ADW55570	ADW55570 Human COL
42	383.5	40.3	1838	7	ADW55520	ADW55520 Human COL
43	383.5	40.3	1838	9	ADW70238	ADW70238 Tumor-aa
44	383.5	40.3	1838	9	ADW70238	ADW70238 Tumor-aa
45	383.5	40.3	1838	9	ADW70492	ADW70492 Human COL
46	382.5	40.1	638	8	ADW08295	ADW08295 Human COL
47	381.5	40.1	703	6	ABP96315	ABP96315 Human COL
48	381.5	40.1	703	6	ADW98189	ADW98189 Protein f
49	381.5	40.1	717	6	ABP96314	ABP96314 Human COL
50	381.5	40.1	733	6	ABU69145	ABU69145 Human COL
51	381.5	40.1	733	6	ADW08293	ADW08293 Human COL
52	381.5	40.1	1181	8	ADP22968	ADP22968 PRO polyp
53	380.5	40.0	638	6	ABU69146	ABU69146 Human COL
54	380	39.9	1767	8	ADQ39813	ADQ39813 Human COL
55	380	39.9	1767	8	ADQ39817	ADQ39817 Human COL
56	380	39.9	1806	8	ADQ39816	ADQ39816 Human COL
57	380	39.9	1806	8	ADQ39815	ADQ39815 Human COL
58	380	39.9	1818	8	ADQ39812	ADQ39812 Human COL
59	380	39.9	1818	8	ADQ39814	ADQ39814 Human COL
60	378.5	39.8	635	4	AAW78798	AAW78798 Human COL
61	378.5	39.8	635	4	AAW78798	AAW78798 Human COL
62	378.5	39.8	644	4	AAW79782	AAW79782 Human COL
63	378.5	39.8	644	4	ADW98659	ADW98659 Protein f
64	378.5	39.8	1745	4	AAW7793	AAW7793 Human COL
65	378.5	39.8	1745	4	ABW97234	ABW97234 Novel hum
66	378.5	39.8	1745	4	ADQ19841	ADQ19841 Human COL
67	377.5	39.7	1466	4	AAE02537	AAE02537 Porcine a
68	377	39.6	1028	8	ADN35278	ADN35278 Helical d
69	377	39.6	1078	2	AAW71704	AAW71704 Collagen
70	377	39.6	1078	3	AAW96125	AAW96125 Collagen
71	377	39.6	1078	5	AAE16478	AAE16478 Human COL
72	377	39.6	1078	5	ABW80736	ABW80736 Collagen
73	377	39.6	1078	5	ABW809628	ABW809628 Amino aci
74	377	39.6	1078	7	ADP13078	ADP13078 Human COL
75	377	39.6	1313	8	ADN35279	ADN35279 Synthetic
76	377	39.6	1313	8	ADN35277	ADN35277 Helical d
77	377	39.6	1449	4	AAE02535	AAE02535 Porcine a
78	377	39.6	1466	4	ABW50291	ABW50291 Collagen
79	377	39.6	1466	5	ABW90747	ABW90747 Human COL
80	377	39.6	1466	5	ABW54454	ABW54454 Human COL
81	377	39.6	1466	6	ABR47418	ABR47418 Breast ca
82	377	39.6	1466	7	ADP65248	ADP65248 Human COL
83	377	39.6	1466	7	ADP65210	ADP65210 Human COL
84	377	39.6	1466	8	ADQ26091	ADQ26091 Type III,
85	377	39.6	1466	8	ADQ26677	ADQ26677 Human COL
86	377	39.6	1466	8	ADP16802	ADP16802 Human COL
87	377	39.6	1466	8	ADP16427	ADP16427 Human COL
88	377	39.6	1466	8	ABW80366	ABW80366 Tumor-aa
89	377	39.6	1466	8	ADW67267	ADW67267 Human COL
90	377	39.6	1466	9	ADW70235	ADW70235 Tumor-aa
91	377	39.6	1466	9	ADZ09873	ADZ09873 Human COL
92	377	39.6	1466	9	AAW04495	AAW04495 Human COL
93	377	39.6	1469	7	ABG15191	ABG15191 Novel hum
94	377	39.6	1470	4	ADW09339	ADW09339 Novel pro
95	377	39.6	1572	8	ADU04510	ADU04510 SPTI-COL
96	377	39.6	1604	8	ADU04497	ADU04497 Modified
97	377	39.6	1726	6	ABR42661	ABR42661 Decorin-m

98	377	39.6	1950	8	ADU04493	AdU04493	Modified
99	375.5	39.4	231	2	AAY23937	Aay23937	Amino aci
100	374	39.3	1518	4	ABG22679	Abg22679	Novel hum

ALIGNMENTS

RESULT 1

ABR48481

ID ABR48481 standard; protein; 163 AA.

XX ABR48481;

XX 13-JUN-2003 (first entry)

XX Human Alpha 1 type XVI collagen.

XX Human; GENSET; therapeutic; therapy.

XX Homo sapiens.

XX WO200294864-A2.

XX 28-NOV-2002.

XX 06-AUG-2001; 2001WO-1B001715.

XX 25-MAY-2001; 2001US-0293574P.

XX 15-JUN-2001; 2001US-0298698P.

XX 29-JUN-2001; 2001US-0302277P.

XX 13-JUL-2001; 2001US-0305456P.

XX (GENSET) GENSET.

XX Bejanin S, Tanaka H;

XX WPI; 2003-129412/12.

XX N-PSDB; ACC51088.

XX New GENSET polynucleotides and polypeptides, useful for preparing a

XX composition for treating GENSET-related disorders and as reagents in

XX assays to quantitatively determined levels of GENSET expression in

XX biological samples.

XX Claim 2; Page 451; 505pp; English.

XX The present invention relates to novel human GENSET coding sequences

XX (ACC51088-ACC51115) and proteins (ABR48453-ABR48508). The GENSET

XX sequences are useful for preparing a composition for treating GENSET-

XX related disorders. They can also be used as markers for tissues in which

XX the corresponding protein is preferentially expressed, as molecular

XX weight markers on Southern gels, as chromosome markers or tags to

XX identify chromosomes, and as reagents in assays to quantitatively

XX determined levels of GENSET expression in biological samples

XX Sequence 163 AA;

XX Query Match 100.0%; Score 952; DB 6; Length 163;

XX Best Local Similarity 100.0%; Pred. No. 1.3e-63; Mismatches 0; Indels 0; Gaps 0;

XX Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ADQ21405	ADQ21405	ADQ21405	ADQ21405	ADQ21405	ADQ21405	ADQ21405	ADQ21405
ID	ADQ21405	standard;	protein;	1603	AA.		

XX ADQ21405;

XX 26-AUG-2004 (first entry)

XX Human soft tissue sarcoma-upregulated protein - SEQ ID 4225.

XX soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human.

XX Homo sapiens.

XX WO2004048938-A2.

XX 10-JUN-2004.

XX 26-NOV-2003; 2003WO-US038193.

XX 26-NOV-2002; 2002US-0429739P.

XX (PROT-) PROTEIN DESIGN LABS INC.

XX Aziz N, Ginsburg WM, Zlotnick A;

XX WPI; 2004-441208/41.

XX Early detection of soft tissue sarcoma comprises determining expression

XX of a gene in a first soft tissue sample and a normal soft tissue sample

XX and comparing the gene expression, also useful in treating soft tissue

XX sarcoma.

XX Example 2; SEQ ID NO 4225; 210pp; English.

XX The invention relates to a novel method for detecting soft tissue sarcoma

XX which comprises obtaining a first soft tissue sample from an individual

XX and a normal soft tissue sample from the same or different individual,

XX determining the expression of a gene in both samples and comparing the

XX expression of the gene in both soft tissue samples, where a higher level

XX of protein expression in the first soft tissue sample indicates the

XX presence of soft tissue sarcoma. The method of the invention has

XX cytosolic applications and may be useful for detecting soft tissue

XX sarcoma, possibly via gene therapy or vaccine production. The nucleic

XX acid sequences may be useful in diagnostic and screening applications.

XX The current sequence is that of a human soft tissue sarcoma-upregulated

XX protein of the invention. The current sequence is not shown within the

XX specification per se but was submitted in CD format by the inventor.

XX Sequence 1603 AA;

XX Query Match 88.3%; Score 840.5; DB 8; Length 1603;

XX Best Local Similarity 44.5%; Pred. No. 2.2e-54; Mismatches 0; Indels 203; Gaps 1;

XX Matches 163; Conservative 0; Mismatches 0; Indels 203; Gaps 1;

```
Qy 93 -----PGPFG 97
Db 1478 GXDQAPRPGAPSPGLPGQIGREGRGQLPGVIRGLPCTKXGKEDIGIGIANGENLPGPFG 1537
Qy 98 PGPFGYGRKAGATGPMQOQIGIPGPFGPMGPGKAGHCNPSDCCFAMPMEOQYPPMKT 157
Db 1538 PGPFGYGRKAGATGPMQOQIGIPGPFGPMGPGKAGHCNPSDCCFAMPMEOQYPPMKT 1597
Qy 158 MKGPGF 163
Db 1598 MKGPGF 1603

RESULT 3
ADL25758
ID ADL25758 standard; protein; 218 AA.
XX
AC ADL25758;
XX
DT 20-MAY-2004 (first entry)
XX
DE Human cancer suppressing protein #6.
XX
KW human; cancer suppression; cancer.
XX
OS Homo sapiens.
XX
PN CN1403478-A.
XX
PD 19-MAR-2003.
XX
PF 12-SEP-2001; 2001CN-00126726.
XX
PR 12-SEP-2001; 2001CN-00126726.
XX
PA (SHAN-) SHANGHAI XINSHIJI GENE TECHN DEV CO LTD.
XX
PI Gu J, Yang S;
XX
DR WPI; 2003-494227/47.
XX
DR N-PSDB; ADL25757, ADL25759.
XX
PT Human protein with function of suppressing cancer cell growth and its
PT coding sequence.
PS Claim 1; SEQ ID NO 17; 43pp; Chinese.
XX
XX The invention relates to one kind of human protein with cancer
XX suppressing function, polynucleotides encoding the polypeptide and the
XX recombinant process of producing the polypeptide. The present invention
XX also discloses the method of using the polypeptide in treating various
XX diseases, such as cancer. The present invention also discloses the
XX agonist resisting the polypeptide and its treatment effect. The present
XX invention also discloses the application of the polynucleotides encoding
XX the human protein with cancer suppressing function. The present sequence
XX represents the amino acid sequence of a human cancer suppressing protein.
SQ Sequence 218 AA;

Query Match 65.1%; Score 620; DB 7; Length 218;
Best Local Similarity 58.0%; Pred. No. 1e-38;
Matches 119; Conservative 7; Mismatches 37; Indels 42; Gaps 3;

Qy 1 MGPPRGKGTGNGHGLPGPKDDCGKPPPGSTG----- 32
Db 14 VGPPRPGRGGHGGAPSPGSPGLPGVPSGMDVMNYDEIKRFRORIIIMFDERMAYYT 73
Qy 33 -----RPGABEGPGAMGPGRPGPHYGPFGPGGPGAGISAV-GLKGDGGA 80
Db 74 SRNQPFMEMMAAARERPEPPKDKGAPGRBGAAGSGGLGQIGREBRQGLPGVRGLPGTKGE 133
Qy 81 TGR--RGLAGLPGQPGPGPGPGGYGKMGATGPMQOQIGIPGPFGPMGPGKAGHCN 138
Db 138 TGR--RGLAGLPGQPGPGPGPGGYGKMGATGPMQOQIGIPGPFGPMGPGKAGHCN 138
```

```
Db 134 KIDIGIGIANGENLPGPFGPGPGYGRKAGATGPMQOQIGIPGPFGPMGPGKAGHCN 193
Qy 139 PSDCFGAMPMEQOYPPMKTMKGPGF 163
Db 194 PSDCFGAMPMEQOYPPMKTMKGPGF 218

RESULT 4
AAM57645
ID AAM57645 standard; peptide; 330 AA.
XX
AC AAM57645;
XX
DT 27-AUG-1998 (first entry)
XX
DE Collagen-like polymer.
XX
KW Collagen-like polymer; synthetic polymer; fibre coating;
KW prosthetic device; catalytic substance.
XX
OS Synthetic.
XX
PN US5773249-A.
XX
PD 30-JUN-1998.
XX
PF 02-MAY-1996; 96US-00642255.
XX
PR 04-NOV-1986; 86US-00927258.
XX
PR 29-OCT-1987; 87US-00114618.
XX
PR 09-NOV-1988; 88US-00269429.
XX
PR 06-NOV-1990; 90US-00609716.
XX
PR 12-NOV-1991; 91US-00791960.
XX
PR 05-NOV-1992; 92US-00972032.
XX
PR 22-DEC-1995; 95US-00577046.
XX
PA (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.
XX
PI Ferrari FA, Cappello J;
XX
DR WPI; 1998-387004/33.
XX
PT Recombinant collagen-like polymers - useful for making gels, films,
PT fibres, etc.
PS Example 2; Col 23; 93pp; English.
XX
XX This sequence represents a unnatural collagen-like polymer of the
XX invention. The products may be used as films, fibres, moulded objects and
XX admixed with other natural or synthetic polymers or coatings on fibres,
XX films, labware or other surfaces, e.g. prosthetic devices. The polymers
XX may be used for binding a wide variety of specific binding materials, as
XX catalytic substances (where the amino acid sequence may specifically
XX chelate a wide variety of elements), as purification media, composites,
XX laminates or adhesives. They may also be combined with inorganic or
XX organic materials such as carbon fibres, nylon fibres, nitrocellulose,
XX etc., as flask coatings or in synthetic matrices for the growth and study
XX of cells, as affinity columns or as supports for biological materials.
XX The polymers have collagen-like properties, but may be easily expressed
XX in micro-organisms in high efficiency. The new sequences can be tailored
XX to give the desired properties
SQ Sequence 330 AA;

Query Match 44.0%; Score 419; DB 2; Length 330;
Best Local Similarity 54.2%; Pred. No. 1.5e-23;
Matches 83; Conservative 2; Mismatches 60; Indels 8; Gaps 2;

Qy 2 GPPGFKGTGTHPLPGPKDDCGKPPPGSTGRCGAGSRGAMGPGRPPRGPHVGPFGP 61
Db 157 GPPGPPGPPAPRPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 216
Qy 62 GQPGPAG-ISAVALKQDRGATGIRGLAGLPGQPGPPPGQPPG-----YGRKAGATGPM 113
```

Dd		217	GPDPAGVGSPPGAPRPPPGPPPCAPPDPGGPAGPVGSFGARPPCP	276
Oy		114	GQDGIPIGPDPGWGQPKXAGHGNPSDCFGAM	146
Dd		277	GPCGPAPGAPRPPGPDPGGPAGPVGSFGAM	309
RESULT 5				
AAW07539	ID			
AAW07539	standard; protein;	408 AA.		
XX	AAW07539;			
XX	25-MAR-2003	(revised)		
DT	03-FEB-1997	(first entry)		
DT				
XX				
DE	Collagen like proteIn (CLP).			
KM	Polymer; repeat unit; natural collagen; intervening oligopeptide; fibre;			
KM	film; membrane; emulsion; coating; collagen like protein;			
KW	specific binding material; catalyst; purification agent; composite;			
KW	laminates; adhesive; cell growth surface; affinity column;			
KW	biological material support; wound dressing; in vivo prothesis.			
XX				
OS	Synthetic.			
XX				
PN	US5514581-A.			
PD				
XX	07-MAY-1996.			
PF				
XX	06-NOV-1990;	90US-00609716.		
PR				
XX	04-NOV-1986;	86US-00927258.		
PR	29-OCT-1987;	87US-00114618.		
PR	09-NOV-1988;	88US-00269429.		
PR	07-NOV-1989;	89WO-US005016.		
XX				
PA	(PROT-) PROTEIN POLYMER TECHNOLOGIES INC.			
PI	Cappello J, Ferrari FA:			
XX				
XX	WPI; 1996-238772/24.			
DR				
PT	DNA encoding protein contg. repeated fibroin derived segments - linked by			
PT	oligopeptide with cell adhesion properties useful, e.g. in wound			
PT	dressings.			
XX				
XX	Example 3; Col 99-102; 71pp; English.			
XX				
CC	Novel DNA sequence encodes a polymer comprising segments of repeating			
CC	units of 3-9 amino acids from natural collagen, i.e. the present sequence			
CC	collagen like protein (CLP), able to assemble into aligned structures			
CC	formable into articles. The polymer comprises at least 2 segments joined			
CC	by an unaligned intervening oligopeptide, other than the repeating unit.			
CC	The polymer can be used to make fibres, films, membranes, emulsions,			
CC	coatings, etc., useful as, e.g. specific binding materials, catalysts,			
CC	purification agents, composites, laminates, adhesives, cell growth surfaces,			
CC	affinity columns and supports for biological materials. Typical			
CC	applications include wound dressings, and in vivo protheses. The polymer			
CC	produces articles with good mechanical properties, and the intervening			
CC	oligopeptide can provide a ligand for binding a mol., antibody, etc., or			
CC	a chemically reactive site for coupling to proteins, etc.. (Updated on 25-			
CC	-MAR-2003 to correct PF field.)			
SO				
Sequence	408 AA;			
Query Match	44.0%; Score 419; DB 2; Length 408;			
Best Local Similarity	54.2%; Pred. No. 1,8e-23;			
Matches	83; Conservative 2; Mismatches 60; Indels 8; Gaps 2;			
Oy		2	GPDPKGTGHPLGLFGRGDGCKPGRPPSGTRPGAEGBGAMGPGRRGPPGHVGPDP	61

Dd	235	GPPGPGGGPAGBPGRPPGPGRPPRPAQVSGAPGRRPGPPRPPAQPGRPPGPP	294
Qy	62	GQPGDAG-ISAVALKGRDGAATGERGLAQPGCGPPGDPGPG-----YGMKATGPM	113
Dd	225	GPPGAGAVSFGAGBPGRPPGPBPGRPPGAPGPPGPPGPPGAPVSGAPGPPGPP	354
Qy	114	GQQGIPIGIPRPPGPMGQPKAKHCPSPDCFAM	146
Dd	355	GPPGPPGAPGPPGPPGPPGPPGPPGAPVSGPM	387
 RESULT 6 AAB64007			
ID	AAB64007	standard; protein; 408 AA.	
AC	AAB64007;		
XX			
DT	19-MAR-2001	(first entry)	
XX			
DE	CLP protein sequence SEQ ID 65.		
KW	proteinaceous polymer; repeat unit; structural polymer; coating; film;		
KX	fibre; membrane; adhesive; emulsion; laminate; keratin; collagen.		
XX			
OS	Synthetic.		
PN	US6140072-A.		
XX			
PD	31-OCT-2000.		
PF	07-JUN-1995;	95US-00475411.	
XX			
PR	04-NOV-1986;	86US-00927258.	
PR	29-OCT-1987;	87US-00114618.	
PR	09-NOV-1988;	88US-00269429.	
PR	06-NOV-1990;	90US-00609716.	
XX			
PA	(PROT-) PROTEIN POLYMER TECHNOLOGIES INC.		
XX			
PI	Cappelletto J, Ferrari FA;		
DR	WPI; 2001-048958/06.		
PT	New DNA encoding a polymer with strands of repeating units of natural		
PT	protein joined by intervening oligopeptide for producing high molecular		
PT	weight polymers of amino acids.		
XX			
PS	Example 2; Col 41; 73pp; English.		
XX			
CC	This invention relates to DNA encoding a proteinaceous polymer. The		
CC	polymer comprises strands of repeating units of a natural protein capable		
CC	of assembling into aligned structures, with at least 2 strands joined by		
CC	an intervening oligopeptide other than the repeating units. The		
CC	intervening oligopeptide is unaligned and the polymer has individual		
CC	strands of the same or different repeating unit. The DNA is useful for		
CC	producing high molecular weight polymers of amino acids based on		
CC	biologically and chemically active structural polymers. These polymers		
CC	may be used to provide a variety of structures for different purposes,		
CC	and to produce articles including coatings, or other (non)structural		
CC	components, e.g. fibres, films, membranes, adhesives or emulsions, or		
CC	with other compounds and/or compositions to form composites or laminates.		
CC	Peptide sequences AAB63371-863391 represent monomer sequences which can		
CC	be used in the polymers of the invention. Oligonucleotide sequences		
CC	AAF23370 - AAF23386 and amino acid sequences AAB6392 - AAB6402 are used		
CC	in the construction of SLP and FCB-SLP polymers. Oligonucleotide		
CC	sequences AAF23387 - AAF23397 and amino acid sequences AAB6403 -		
CC	AAB6408 are used in the construction of CLP (collagen like protein)		
CC	polymers. Oligonucleotide sequences AAF23398 - AAF23409 and amino acid		
CC	sequences AAB6409 - AAB64014 are used in the construction of keratin		
CC	polymers. Proteins and peptides represented by sequences AAB64015 -		
CC	AAB64049 are examples of polymers of the invention		
XX			
SQ	Sequence 408 AA;		

Query Match 44.0%; Score 419; DB 4; Length 408;
 Best Local Similarity 54.2%; Pred. No. 1.8e-23;
 Matches ... 83; Conservative 2; Mismatches 60; Indels 8; Gaps 2;

QY 2 GPPGFKGKTGHPGLPGRPKDCGKRPBGSTGRPGAGEPGAMKPGRPGHPGVSPGPP 61
 DB 235 GPPGPPGPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 294
 QY 62 GQGPAG-1SAVGLKDRGATGERGLAGLPQGPBPBGPPG-----YGMGATGPM 113
 DB 295 GPPGPPGPPGSPAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 354
 QY 114 GQGIPIGIPGPPPMGQPGKAGHCNPSDCFGAM 146
 DB 355 GPPGPPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 387

RESULT 7
 AAB72737
 ID AAB72737 standard; protein; 408 AA.
 AC AAB72737;
 XX
 XX
 DT 09-MAY-2001 (first entry)
 DE Repetitive protein polymer protein sequence #37.
 XX
 XX Protein polymer; repetitive sequence; repeat; mechanical property;
 KW coating; fibre; film; membrane; adhesive; emulsion; composite; laminate;
 KM diagnosis; sensor.
 XX
 OS Unidentified.
 XX
 XX US6184348-B1.
 PN 06-FEB-2001.
 PD 07-JUN-1995; 95US-00478029.
 XX
 PF 04-NOV-1986; 86US-00927258.
 PR 29-OCT-1987; 87US-00114618.
 PR 09-NOV-1988; 88US-00269429.
 PR 06-NOV-1990; 90US-00609716.
 XX
 PA (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.
 XX
 PI Ferrari FA, Cappello J;
 XX
 DR WPI; 2001-217800/22.
 XX
 XX New recombinant proteinaceous polymers comprising strands of repeating
 PT units of natural protein which can be assembled into aligned structures,
 PT useful for producing e.g. fibers, films, membranes, adhesives, or
 PT emulsions.
 PS Example 3; Col 99-102; 72pp; English.
 XX
 XX The present invention provides compositions containing proteinaceous
 CC polymers comprising proteins with strands of repeating units of a natural
 CC protein. These can be used as coatings, fibres, films, membranes,
 CC adhesives, emulsions, in composites and laminates, and in affinity
 CC columns, diagnostic devices and sensors. The present sequence is a
 CC protein sequence used as a demonstration of the invention
 XX
 SQ Sequence 408 AA;

Query Match 44.0%; Score 419; DB 4; Length 408;
 Best Local Similarity 54.2%; Pred. No. 1.8e-23;
 Matches 83; Conservative 2; Mismatches 60; Indels 8; Gaps 2;

QY 2 GPPGFKGKTGHPGLPGRPKDCGKRPBGSTGRPGAGEPGAMKPGRPGHPGVSPGPP 61
 DB 235 GPPGPPGPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 294
 QY 62 GQGPAG-1SAVGLKDRGATGERGLAGLPQGPBPBGPPG-----YGMGATGPM 113
 DB 295 GPPGPPGPPGSPAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 354
 QY 114 GQGIPIGIPGPPPMGQPGKAGHCNPSDCFGAM 146
 DB 355 GPPGPPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 387

DB 235 GPPGPPGPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 294
 QY 62 GQGPAG-1SAVGLKDRGATGERGLAGLPQGPBPBGPPG-----YGMGATGPM 113
 DB 295 GPPGPPGPPGSPAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 354
 QY 114 GQGIPIGIPGPPPMGQPGKAGHCNPSDCFGAM 146
 DB 355 GPPGPPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 387

RESULT 8
 AAR95115
 ID AAR95115 standard; protein; 357 AA.
 AC AAR95115;
 XX
 XX
 DT 25-MAR-2003 (revised)
 DT 03-FEB-1997 (first entry)
 DE Intervening sequence contg. collagen like protein (CLP)-CB.
 XX
 XX Polymer; repeat unit; natural collagen; intervening oligopeptide; fibre;
 KW film; membrane; emulsion; coating; collagen like protein;
 KW specific binding material; catalyst; purification agent; composite;
 KM laminate; adhesive; cell growth surface; affinity column;
 XX biological material support; wound dressing; in vivo prothesis.
 OS Synthetic.
 XX
 XX US5514581-A.
 PN 07-MAY-1996.
 PD 06-NOV-1990; 90US-00609716.
 XX
 PF 04-NOV-1986; 86US-00927258.
 PR 29-OCT-1987; 87US-00114618.
 PR 09-NOV-1988; 88US-00269429.
 PR 07-NOV-1989; 89WO-US005016.
 XX
 PA (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.
 XX
 PI Cappello J, Ferrari FA;
 XX
 DR WPI; 1996-238772/24.
 XX
 XX DNA encoding protein contg. repeated fibroin derived segments - linked by
 PT oligopeptide with cell adhesion properties useful, e.g. in wound
 PT dressings.
 PS Example 3; Col 101-104; 71pp; English.
 XX
 XX Novel DNA sequence encodes a polymer comprising segments of repeating
 CC units of 3-9 amino acids from natural collagen, i.e. the present
 CC intervening sequence contg. collagen like protein (CLP)-CB, able to
 CC assemble into aligned structures formable into articles. The polymer
 CC comprises at least 2 segments joined by an unaligned intervening
 CC oligopeptide, other than the repeating unit. The polymer can be used to
 CC make fibres, films, membranes, emulsions, coatings, etc., useful as, e.g.
 CC specific binding materials, catalysts, purificn. agents, composites,
 CC laminates, adhesives, cell growth surfaces, affinity columns and supports
 CC for biological materials. Typical applications include wound dressings,
 CC and in vivo protheses. The polymer produces articles with good mechanical
 CC properties, and the intervening oligopeptide can provide a ligand for
 CC binding a mol., antibody, etc., or a chemically reactive site for
 CC coupling to proteins, etc. (Updated on 25-MAR-2003 to correct PF field.)
 XX
 SQ Sequence 357 AA;

Query Match 43.3%; Score 412.5; DB 2; Length 357;
 Best Local Similarity 42.7%; Pred. No. 4.9e-23;
 Matches 90; Conservative 6; Mismatches 54; Indels 61; Gaps 6;

CC with function superiority to the native collagen.

XX Sequence 1071 AA;

Query Match 42.0%; Score 399.5; DB 7; Length 1071;
Best Local Similarity 48.9%; Pred. No. 1.2e-21;
Matches 88; Conservative 1; Mismatches 68; Indels 23; Gaps 6;

QY 2 GPPGFKGKTGHP---GLPQPKDCKPQPGSTGRPGAGEFGAMGP---QGRPGPGHV 55
DB 865 GPPGAPGAPGPGSAGAPGPPGAPGPPGAGAPGPPGAGAPGPPGAGAPGPPGSA 924
QY 56 GPPGPPGQPPGAPG---ISAVGLKDDRGATGERTGLAGLPQGPQPPGQPPGV----- 104
DB 925 GAPGPPGAPGAPGPPGAGAPGPPGAPGPPGAGAPGPPGAPGPPGAGAPGPP 984
QY 105 GKMGTATGPMGQGGIGIPGPPPMGQPGKAGHCNPSDCFG-AMPMEQYPPMKTKMGPPG 163
DB 985 GAPGAPGPPGSAAGAPGPPGAPGPPGAGAPGPPGAGAPGPPGAGAPGPPGAGP-----AGPLG 1039

RESULT 13

ABR01788
ID ABR01788 standard; protein; 67 AA.

XX ABR01788;

DT 22-APR-2003 (first entry)

XX Human breast specific polypeptide #103.

XX Human; breast specific nucleic acid; BSNM; breast; cytostatic;
KW gene therapy; vaccines; lung cancer; breast cancer;
KM breast specific polypeptide; BSP.

XX Homo sapiens.

XX WO200268645-A2.

XX 06-SEP-2002.

XX 20-NOV-2001; 2001WO-US045151.

XX 20-NOV-2000; 2000US-024992P.

XX (DIAD-) DIADEXUS INC.

PI Salceda S, Macina RA, Recipon H, Cafferkey R, Sun Y, Liu C;
PI Turner LR;

XX WPI; 2002-713379/77.

PT New breast specific genes and proteins, useful in gene therapy or as
PT vaccines for treating breast cancer or non-cancerous breast diseases, as
PT well as for diagnosing, monitoring or staging these diseases.

XX Claim 11; Page 277; 277p; English.

CC The invention relates to a novel isolated breast specific nucleic acid
CC molecule. The polypeptides of the invention have cytostatic activity. The
CC as novel nucleic acids and polypeptides may have a use in gene therapy, and
CC as vaccines. The breast specific nucleic acid and polypeptide are useful
CC for diagnosing and monitoring the presence and metastases of lung cancer
CC in a patient. The antibody that specifically binds to the breast specific
CC polypeptide is useful for determining the presence of a breast specific
CC protein in a sample, as well as for treating a patient with breast
CC cancer, particularly by inducing an immune response against the breast
CC cancer cell expressing the breast specific nucleic acid molecule or
CC polypeptide. In particular, these breast specific genes and proteins are
CC useful for identifying, diagnosing, monitoring, staging, imaging and
CC treating breast cancer and non-cancerous disease states in breast tissue.
CC These are also useful in gene therapy, production of transgenic animals
CC and cells, and in the production of engineered breast tissue for

CC treatment and research. The sequences shown in ABR01686-ABR01788
CC represent the novel human breast specific polypeptides of the invention

XX Sequence 67 AA;

Query Match 41.5%; Score 395; DB 5; Length 67;
Best Local Similarity 100.0%; Pred. No. 2.2e-22;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 97 GPGPPGYGKMGATGPMGQGGIGIPGPPPMGQPGKAGHCNPSDCFGAMPMEQYPPMK 156
DB 1 GPGPPGYGKMGATGPMGQGGIGIPGPPPMGQPGKAGHCNPSDCFGAMPMEQYPPMK 60
QY 157 TMKGPPG 163
DB 61 TMKGPPG 67

RESULT 14

ADV95808
ID ADV95808 standard; protein; 503 AA.

XX ADV95808;

DT 10-MAR-2005 (first entry)

XX Chicken type IX collagen alpha 1 precursor.

XX arthritis; collagen type IX alpha 1 chain; antiarthritis;
KW antiinflammatory; immunosuppressive; musculoskeletal disease;
KM degeneration.

XX Gallus sp.

XX WO2004110475-A1.

XX 23-DEC-2004.

XX 17-JUN-2004; 2004WO-AU000788.

XX 17-JUN-2003; 2003AU-00903037.

XX (NUTR-) INST NUTRACEUTICAL RES PTY LTD.

XX Ghosh P;

XX WPI; 2005-039981/04.

PT Pharmaceutical composition for treating or preventing arthritis or other
PT degenerative disease in an individual, comprises a polypeptide comprising
PT a collagen type IX alpha 1 chain NC4 domain.

PS Disclosure; SEQ ID NO 17; 92p; English.

CC This invention describes a novel pharmaceutical composition for treating
CC or preventing arthritis or other degenerative disease which comprises a
CC polypeptide a collagen type IX alpha 1 chain NC4 domain or a biologically
CC active fragment having antiarthritis, antiinflammatory and
CC immunosuppressive activity in combination with a carrier. The invention
CC describes two methods for recovering a polypeptide having anti-arthritis
CC or anti-inflammatory activity. The first method comprises isolating a
CC mixture comprising a GAG-peptide and a polypeptide having a molecular
CC weight of less than 30000 Da by autolysis from connective tissue,
CC separating the GAG-peptide from the polypeptide, and recovering the
CC polypeptide. The second method comprises incubating connective tissue in
CC an autolysis medium that provides a buffered pH range of 2.5-8.5 for a
CC time and under conditions sufficient to release a GAG-peptide and a
CC polypeptide having a molecular weight of less than 30000 Da, recovering a
CC mixture comprising the GAG-peptide and polypeptide from the autolysis
CC medium, separating the polypeptide from the GAG-peptide, and recovering
CC the polypeptide having a molecular weight of less than 30000 Da. The
CC recovered anti-arthritis or anti-inflammatory polypeptides are useful for
CC inducing cartilage formation or for preparing a medicament for the

THIS PAGE BLANK (USP 10)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 1, 2005, 08:19:37 ; Search time 46 Seconds
(without alignments)
292.959 Million cell updates/sec

Title: US-09-924-340-58
Perfect score: 952
Sequence: 1 MCPPGFKGTHGHPGLGPKG.....GAMPMEQOYPMKTKMGPRG 163

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/1/iaa/5.COMB.pep:*
2: /cgn2_6/prodata/1/iaa/6.COMB.pep:*
3: /cgn2_6/prodata/1/iaa/H.COMB.pep:*
4: /cgn2_6/prodata/1/iaa/PCTUS.COMB.pep:*
5: /cgn2_6/prodata/1/iaa/RE.COMB.pep:*
6: /cgn2_6/prodata/1/iaa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	952	100.0	163	2	US-10-000-489-58	Sequence 58, App1
2	840.5	88.3	1603	2	US-09-949-016-6136	Sequence 6136, Ap
3	840.5	88.3	1609	2	US-09-949-016-10910	Sequence 10910, A
4	419	44.0	330	1	US-08-642-255-32	Sequence 32, App1
5	419	44.0	408	1	US-07-609-716-65	Sequence 65, App1
6	419	44.0	408	2	US-08-475-411A-65	Sequence 65, App1
7	419	44.0	408	2	US-08-478-029A-65	Sequence 65, App1
8	412.5	43.3	357	1	US-07-609-716-66	Sequence 66, App1
9	412.5	43.3	357	1	US-08-642-255-33	Sequence 66, App1
10	412.5	43.3	357	2	US-08-475-411A-66	Sequence 66, App1
11	412.5	43.3	357	2	US-08-478-029A-66	Sequence 66, App1
12	390	41.0	1268	2	US-08-494-168-10	Sequence 10, App1
13	390	41.0	1268	2	US-09-949-016-7487	Sequence 7487, Ap
14	389.5	40.9	1064	1	US-08-642-255-62	Sequence 62, App1
15	388.5	40.8	504	2	US-09-219-849-3	Sequence 3, App1
16	388.5	40.8	551	1	US-08-642-255-52	Sequence 52, App1
17	386	40.5	252	1	US-08-642-255-61	Sequence 61, App1
18	385	40.4	234	1	US-08-642-255-51	Sequence 51, App1
19	378.5	39.8	638	2	US-10-001-887-108	Sequence 108, App
20	378.5	39.8	1745	2	US-09-795-061-4	Sequence 4, App1
21	378.5	39.8	1745	2	US-09-949-002-405	Sequence 405, App
22	378.5	39.8	1771	2	US-09-949-002-492	Sequence 492, App
23	377	39.6	1057	2	US-08-931-820-4	Sequence 4, App1
24	377	39.6	1078	2	US-08-963-825-21	Sequence 21, App1
25	377	39.6	1078	2	US-09-500-811-21	Sequence 21, App1
26	377	39.6	1078	2	US-09-570-573-21	Sequence 21, App1
27	377	39.6	1078	2	US-09-548-608-21	Sequence 21, App1

28	376.5	39.5	1690	2	US-09-949-016-5884	Sequence 5884, Ap
29	372.5	39.1	417	1	US-08-175-155-69	Sequence 69, App1
30	372.5	39.1	417	1	US-08-477-509B-104	Sequence 104, App
31	372.5	39.1	417	1	US-08-642-255-102	Sequence 102, App
32	372.5	39.1	417	1	US-08-707-237A-76	Sequence 76, App1
33	372.5	39.1	417	2	US-08-482-085B-104	Sequence 104, App
34	372.5	39.1	417	2	US-09-444-791A-104	Sequence 104, App
35	372.5	39.1	680	2	US-09-949-001-20	Sequence 20, App1
36	372.5	39.1	680	2	US-09-961-403-5	Sequence 5, App1
37	372.5	39.1	720	2	US-09-219-849-4	Sequence 4, App1
38	372.5	39.1	777	1	US-08-642-255-53	Sequence 53, App1
39	372.5	39.1	829	1	US-08-642-255-132	Sequence 132, App
40	372.5	39.1	829	1	US-08-397-633A-53	Sequence 53, App1
41	372.5	39.1	837	1	US-08-175-155-68	Sequence 68, App1
42	372.5	39.1	837	1	US-08-477-509B-103	Sequence 103, App
43	372.5	39.1	837	1	US-08-642-255-101	Sequence 101, App
44	372.5	39.1	837	1	US-08-707-237A-75	Sequence 75, App1
45	372.5	39.1	837	2	US-08-482-085B-103	Sequence 103, App
46	372.5	39.1	837	2	US-09-444-791A-103	Sequence 103, App
47	372.5	39.1	897	2	US-08-397-633A-50	Sequence 50, App1
48	372	39.1	1057	2	US-08-931-820-1	Sequence 1, App1
49	372	39.1	1057	2	US-10-153-469A-16	Sequence 16, App1
50	372	39.1	1057	2	US-10-153-469A-20	Sequence 20, App1
51	372	39.1	1057	2	US-10-104-889-16	Sequence 16, App1
52	372	39.1	1057	2	US-10-104-889-20	Sequence 20, App1
53	372	39.1	1107	2	US-10-153-469A-11	Sequence 11, App1
54	372	39.1	1107	2	US-10-104-889-11	Sequence 11, App1
55	372	39.1	1171	2	US-10-153-469A-8	Sequence 8, App1
56	372	39.1	1171	2	US-10-104-889-8	Sequence 8, App1
57	372	39.1	1218	2	US-09-949-016-7065	Sequence 7065, Ap
58	372	39.1	1388	2	US-10-153-469A-10	Sequence 10, App1
59	372	39.1	1388	2	US-10-104-889-10	Sequence 10, App1
60	372	39.1	1464	2	US-09-331-347-21	Sequence 21, App1
61	371.5	39.0	680	2	US-09-949-001-15	Sequence 15, App1
62	371.5	39.0	1806	2	US-09-919-497-56	Sequence 56, App1
63	371	38.9	1017	2	US-08-468-996-10	Sequence 10, App1
64	370.5	38.9	1442	1	US-08-316-650-12	Sequence 12, App1
65	370.5	38.9	1442	4	PCT-US93-02251-12	Sequence 6, App1
66	370	38.9	1169	2	US-10-104-889-6	Sequence 6, App1
67	370	38.9	1169	2	US-10-104-889-6	Sequence 6, App1
68	370	38.9	1341	2	US-08-963-825-18	Sequence 18, App1
69	370	38.9	1341	2	US-09-500-811-18	Sequence 18, App1
70	370	38.9	1341	2	US-09-570-573-18	Sequence 18, App1
71	370	38.9	1341	2	US-09-548-608-18	Sequence 18, App1
72	370	38.9	1461	2	US-09-585-887-9	Sequence 9, App1
73	370	38.9	1461	2	US-09-289-578-9	Sequence 9, App1
74	369.5	38.8	1060	2	US-08-531-820-3	Sequence 3, App1
75	369.5	38.8	1418	2	US-08-963-825-20	Sequence 20, App1
76	369.5	38.8	1418	2	US-09-10-999-1	Sequence 1, App1
77	369.5	38.8	1418	2	US-09-500-811-20	Sequence 20, App1
78	369.5	38.8	1418	2	US-09-570-573-20	Sequence 20, App1
79	369.5	38.8	1418	2	US-09-548-608-20	Sequence 20, App1
80	369	38.8	532	1	US-08-994-168-9	Sequence 9, App1
81	368	38.7	938	2	US-09-949-016-9992	Sequence 9992, Ap
82	366	38.4	1739	2	US-09-795-061-2	Sequence 2, App1
83	365.5	38.4	492	2	US-08-468-996-12	Sequence 12, App1
84	365.5	38.4	822	2	US-09-219-849-49	Sequence 49, App1
85	365	38.3	689	2	US-09-949-016-6137	Sequence 11276, A
86	362	38.0	595	2	US-09-819-849-48	Sequence 48, App1
87	362	38.0	595	2	US-09-819-849-50	Sequence 50, App1
88	361.5	38.0	1078	2	US-09-949-016-11185	Sequence 11185, A
89	361.5	38.0	1143	2	US-09-949-016-6137	Sequence 6137, Ap
90	360.5	37.9	347	2	US-09-623-497-1	Sequence 1, App1
91	360.5	37.9	623	2	US-09-829-348-3	Sequence 3, App1
92	360.5	37.9	626	2	US-09-829-348-2	Sequence 2, App1
93	360.5	37.9	684	1	US-08-555-669-12	Sequence 12, App1
94	360.5	37.9	684	2	US-09-073-663-13	Sequence 13, App1
95	360.5	37.9	1516	2	US-09-949-016-8209	Sequence 8209, Ap
96	359.5	37.8	219	2	US-10-153-469A-44	Sequence 44, App1
97	359.5	37.8	219	2	US-10-153-469A-46	Sequence 46, App1
98	359.5	37.8	219	2	US-10-104-889-44	Sequence 44, App1
99	359.5	37.8	219	2	US-10-104-889-46	Sequence 46, App1
100	359.5	37.8	234	2	US-09-895-674A-1	Sequence 1, App1

ALIGNMENTS

```
RESULT 1
US-10-000-489-58
; Sequence 58, Application US/10000489
; Patent No. 6794363
; GENERAL INFORMATION:
; APPLICANT: Benjann, Stephanie
; APPLICANT: Tanaka, Hiroaki
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
; FILE REFERENCE: 91.US6.DIV
; CURRENT APPLICATION NUMBER: US/10/000,489
; CURRENT FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: US 09/924,340
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: PCT/IB01/01715
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/305,456
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/302,277
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/298,698
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/293,574
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: JPatent
; SEQ ID NO 58
; LENGTH: 163
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-000-489-58

Query Match      100.0%; Score 952; DB 2; Length 163;
Best Local Similarity 100.0%; Pred. No. 1,1e-65;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGPPGKGTGHPGLPGPKDCKGPPPGSTGRPGAEGEPGAMGPGRPPGHVGP 60
DB 1 MGPPGKGTGHPGLPGPKDCKGPPPGSTGRPGAEGEPGAMGPGRPPGHVGP 60
QY 61 PQGPAGISAVGLKGDRTGATGERGLAGLPGQGPDPGPGQGPYGMGATGPMGQGGIRG 120
DB 61 PQGPAGISAVGLKGDRTGATGERGLAGLPGQGPDPGPGQGPYGMGATGPMGQGGIRG 120
QY 121 IQGPPGPGQPGKAGHCNPSDCGAMPMEQYPPMKTMGPPFG 163
DB 121 IQGPPGPGQPGKAGHCNPSDCGAMPMEQYPPMKTMGPPFG 163

RESULT 2
US-09-949-016-6136
; Sequence 6136, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6136
```

```
; LENGTH: 1603
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6136

Query Match      88.3%; Score 840.5; DB 2; Length 1603;
Best Local Similarity 44.5%; Pred. No. 2.9e-56;
Matches 163; Conservative 0; Mismatches 0; Indels 203; Gaps 1;

QY 1 MGPPGKGTGHPGLPGPKDCKGPPPGSTGRPGAEGEPGAMGPGRPPGHVGP 60
DB 1238 MGPPGKGTGHPGLPGPKDCKGPPPGSTGRPGAEGEPGAMGPGRPPGHVGP 1297
QY 61 PQGPAGISAVGLKGDRTGATGERGLAGLPGQ----- 92
DB 1298 PQGPAGISAVGLKGDRTGATGERGLAGLPGQPPGHPGPPGEPDTGAMGKGGPPGKQ 1357
QY 93 ----- 92
DB 1358 GFYPPGPKDPPGAAQKQAGEKRGAMPGGPKSGSMGPVGPDPAGEGHPGAPGS 1417
QY 93 ----- 92
DB 1418 GSPGLPGVPGSMGMVNYDEIRFLRQELIKMFDERMAVYTSRMQPFMEMAALAPGRPP 1477
QY 93 -----PGPPG 97
DB 1478 GKDAPGRGAPSDGLPQQLIGRGGQLPGVRLPGTKGKGDIGIGIAGENGJLPGPPG 1537
QY 98 PQGPYGMGATGPMGQGGIRGIRGPPGPMGQPGKAGHCNPSDCGAMPMEQYPPMKT 157
DB 1538 PQGPYGMGATGPMGQGGIRGIRGPPGPMGQPGKAGHCNPSDCGAMPMEQYPPMKT 1597
QY 158 MKGPPG 163
DB 1598 MKGPPG 1603

RESULT 3
US-09-949-016-10910
; Sequence 10910, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10910
; LENGTH: 1609
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10910

Query Match      88.3%; Score 840.5; DB 2; Length 1609;
Best Local Similarity 44.5%; Pred. No. 2.9e-56;
Matches 163; Conservative 0; Mismatches 0; Indels 203; Gaps 1;

QY 1 MGPPGKGTGHPGLPGPKDCKGPPPGSTGRPGAEGEPGAMGPGRPPGHVGP 60
DB 1244 MGPPGKGTGHPGLPGPKDCKGPPPGSTGRPGAEGEPGAMGPGRPPGHVGP 1303
QY 61 PQGPAGISAVGLKGDRTGATGERGLAGLPGQ----- 92
```

Db 1304 PGCPGAGISAVGLKGDGATGERGLAGLPQGPGRPHGPPRPGEPGTGDAAGKEGPPKQ 1363
QY 93 ----- 92
Db 1364 GFYGPGRPKDDPGAAGKQAGKAGKAGMPCGPKSGSMKPVGPDPGAGRGHGAAGPS 1423
QY 93 ----- 92
Db 1424 GSPDLPGVPSMGDMVYDEIKRFIRQELIKMFDERMAYYTSMQPPMEMAAAPGRPP 1483
QY 93 -----RPPG 97
Db 1484 GKDAAPRPGAPSSPGLPGQIGRGQLPVRGLPCTKGEKDDIGLAGENGGLPQPPG 1543
QY 98 PGPGPGYKMGATGPMQGGIPGIPRPPPMQPGKAGHCNPSDCFGAMPEDQYPPMKT 157
Db 1544 PGPPPGYKMGATGPMQGGIPGIPRPPPMQPGKAGHCNPSDCFGAMPEDQYPPMKT 1603
QY 158 MKGPPG 163
Db 1604 MKGPPG 1609

RESULT 4
US-08-642-255-32
; Sequence 32, Application US/08642255
; Patent No. 5773249
; GENERAL INFORMATION:
; APPLICANT: CARPELLO, Joseph
; APPLICANT: FERRARI, Franco A.
; TITLE OF INVENTION: High Molecular Weight Collagen-Like
; TITLE OF INVENTION: Protein Polymers
; NUMBER OF SEQUENCES: 135
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/642,255
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: ROWLAND, Berttram I.
; REGISTRATION NUMBER: 20,015
; REFERENCE/DOCKET NUMBER: A5556-3/BIR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8771
; TELEX: 910 277299 FHT UR
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 330 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-642-255-32

Query Match 44.0%; Score 419; DB 1; Length 330;
Best Local Similarity 54.2%; Pred. No. 8, 4e-25;
Matches 83; Conservative 2; Mismatches 60; Indels 8; Gaps 2;
QY 2 GPPGPGYKMGATGPMQGGIPGIPRPPPMQPGKAGHCNPSDCFGAMPEDQYPPMKT 61
Db 157 GPPGPGYKMGATGPMQGGIPGIPRPPPMQPGKAGHCNPSDCFGAMPEDQYPPMKT 216

QY 62 GPPGAG-ISAVGLKGDGATGERGLAGLPQGPGRPHGPPRPGEPGTGDAAGKEGPPKQ 113
Db 217 GPPGAGPVGSPAPRPPRPPGPPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 276
QY 114 GQGGIPGIPRPPPMQPGKAGHCNPSDCFGAM 146
Db 277 GPPGPGAGPPGPPRPPRPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 309

RESULT 5
US-07-609-716-65
; Sequence 65, Application US/07609716
; Patent No. 5514581
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; APPLICANT: Carpello, Joseph
; TITLE OF INVENTION: Functional Recombinantly Prepared
; TITLE OF INVENTION: Synthetic Protein Polymer
; NUMBER OF SEQUENCES: 118
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/609,716
; FILING DATE: 06-NOV-1990
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Berttram I.
; REGISTRATION NUMBER: 20015
; REFERENCE/DOCKET NUMBER: A-55186-3/BIR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 408 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-609-716-65

Query Match 44.0%; Score 419; DB 1; Length 408;
Best Local Similarity 54.2%; Pred. No. 1e-24;
Matches 83; Conservative 2; Mismatches 60; Indels 8; Gaps 2;
QY 2 GPPGPGYKMGATGPMQGGIPGIPRPPPMQPGKAGHCNPSDCFGAMPEDQYPPMKT 61
Db 235 GPPGPGYKMGATGPMQGGIPGIPRPPPMQPGKAGHCNPSDCFGAMPEDQYPPMKT 294
QY 62 GPPGAG-ISAVGLKGDGATGERGLAGLPQGPGRPHGPPRPGEPGTGDAAGKEGPPKQ 113
Db 295 GPPGAGPVGSPAPRPPRPPGPPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 294
QY 114 GQGGIPGIPRPPPMQPGKAGHCNPSDCFGAM 146
Db 355 GPPGPGAGPPGPPRPPRPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 387
RESULT 6
US-08-475-411A-65
; Sequence 65, Application US/08475411A
; Patent No. 6140072

```

; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; APPLICANT: Cappello, Joseph
; TITLE OF INVENTION: Functional Recombinantly Prepared
; TITLE OF INVENTION: Synthetic Protein Polymer
; NUMBER OF SEQUENCES: 119
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,411A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/609,716
; FILING DATE: 06-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/269,429
; FILING DATE: 09-NOV-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/114,618
; FILING DATE: 29-OCT-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/927,258
; FILING DATE: 04-NOV-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: Treccatin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-55186-9/RFT/MTK
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 408 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-475-411A-65

Query Match          44.0%; Score 419; DB 2; Length 408;
Best Local Similarity 54.2%; Pred. No. 1e-24;
Matches 83; Conservative 2; Mismatches 60; Indels 8; Gaps 2;

QY 2 GPPGFGKGTGHTGLPKPKDCKPDPGSTGRPGAEGEGAMPGQRPDPGHVGPDP 61
   |||||
DB 235 GPPGPPGPPGAGPDPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 294
   |||||
QY 62 GPPGPPG-ISAVALGKDRGATGERGLAGLPGQPPGPPGPPGPPGPPGPPGPPG 113
   |||||
DB 295 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 354
   |||||
QY 114 GQGGIPGIPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 146
   |||||
DB 355 GPPGPPGAGPDPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 387
   |||||

RESULT 7
US-08-478-029A-65
; Sequence 65, Application US/08478029A
; Patent No. 6184348
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; TITLE OF INVENTION: Functional Recombinantly Prepared
```

```

; APPLICANT: Cappello, Joseph
; TITLE OF INVENTION: Functional Recombinantly Prepared
; TITLE OF INVENTION: Synthetic Protein Polymer
; NUMBER OF SEQUENCES: 119
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,029A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/609,716
; FILING DATE: 06-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/269,429
; FILING DATE: 09-NOV-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/114,618
; FILING DATE: 29-OCT-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/927,258
; FILING DATE: 04-NOV-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: Treccatin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-55186-8/RFT/MTK
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 408 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-478-029A-65

Query Match          44.0%; Score 419; DB 2; Length 408;
Best Local Similarity 54.2%; Pred. No. 1e-24;
Matches 83; Conservative 2; Mismatches 60; Indels 8; Gaps 2;

QY 2 GPPGFGKGTGHTGLPKPKDCKPDPGSTGRPGAEGEGAMPGQRPDPGHVGPDP 61
   |||||
DB 235 GPPGPPGPPGAGPDPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 294
   |||||
QY 62 GPPGPPG-ISAVALGKDRGATGERGLAGLPGQPPGPPGPPGPPGPPGPPGPPG 113
   |||||
DB 295 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 354
   |||||
QY 114 GQGGIPGIPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 146
   |||||
DB 355 GPPGPPGAGPDPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 387
   |||||

RESULT 8
US-07-609-716-66
; Sequence 66, Application US/07609716
; Patent No. 5514581
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; APPLICANT: Cappello, Joseph
; TITLE OF INVENTION: Functional Recombinantly Prepared
```


QY	161	PEG	163
Db	290	PAG	292

RESULT 15
US-09-219

US-09-219-849-3

; Sequence 3, Application US/09219849

Patent No. 6150081

GENERAL INFORMATION:

APPLICANT: VAN HEERDE, GEORGE V.

APPLICANT: VAN RIJN, ALEXIS C.

APPLICANT: BOUTWSTRA, JAN B.

APPLICANT: DE WOLF, FREDERIK A.

APPLICANT: MOOBROEK, ANDREAS

APPLICANT: WERTEN, MARC W.T.

APPLICANT: WIND, RICHELE D.

APPLICANT: VAN DEN BOSCH, TANJA J.

TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN

TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE

TITLE OF INVENTION: PREPARATION THEREOF

FILE REFERENCE: 2728-2

CURRENT APPLICATION NUMBER: US/09/219,849

CURRENT FILING DATE: 1998-12-23

NUMBER OF SEQ ID NOS: 50

```

; SOFTWARE: PatentIn Ver. 2.1
00000000

```

; SEQ ID NO 3

LENGTH: 504

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: Illustrative

OTHER INFORMATION

US-09-219-849-3

Query Match	40.8%;	Score 388.5;	DB 2;	Length 504;
-------------	--------	--------------	-------	-------------

Best Local Similarity 47.5%; Pred. No. 2.6e-22;

Matches	87;	Conservative	2;	Mismatches	71;	Indels	23;	Gaps	6;
---------	-----	--------------	----	------------	-----	--------	-----	------	----

[illegible]

Search completed: December 1, 2005, 08:28:05
Job time : 53 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 1, 2005, 08:27:13 ; Search time 162 Seconds
(without alignments)
420.409 Million cell updates/sec

Title: US-09-924-340-58

Perfect score: 952
Sequence: 1 MGPMPGKGTGHTGHPGPKG.....GAMPMEQGYPMKTMKGPFG 163

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 100 summaries

Database : Published Applications_AA_Main:*
1: /cgn2_6/prodata/1/pubppaa/US07_PUBCOMB.pep:*
2: /cgn2_6/prodata/1/pubppaa/US08_PUBCOMB.pep:*
3: /cgn2_6/prodata/1/pubppaa/US09_PUBCOMB.pep:*
4: /cgn2_6/prodata/1/pubppaa/US10_PUBCOMB.pep:*
5: /cgn2_6/prodata/1/pubppaa/US10B_PUBCOMB.pep:*
6: /cgn2_6/prodata/1/pubppaa/US11_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	952	100.0	163	3	US-09-992-600A-58 Sequence 58, App1
2	952	100.0	163	3	US-09-924-340-58 Sequence 58, App1
3	952	100.0	163	3	US-09-992-095B-58 Sequence 58, App1
4	952	100.0	163	3	US-09-999-570-58 Sequence 58, App1
5	952	100.0	163	4	US-10-000-489-58 Sequence 58, App1
6	952	100.0	163	4	US-10-000-986-58 Sequence 58, App1
7	952	100.0	163	4	US-10-154-678-58 Sequence 58, App1
8	952	100.0	163	5	US-10-838-854-58 Sequence 58, App1
9	840.5	88.3	1603	5	US-10-723-860-4225 Sequence 4225, App
10	405	42.5	180	5	US-10-797-606-39 Sequence 39, App1
11	395	41.5	67	4	US-10-001-843-218 Sequence 218, App
12	395	41.5	67	6	US-11-005-609-218 Sequence 218, App
13	390	41.0	910	5	US-10-723-860-2221 Sequence 2221, App
14	390	41.0	1693	5	US-10-450-763-35198 Sequence 35198, A
15	390	41.0	1693	5	US-10-450-763-45978 Sequence 45978, A
16	388.5	40.8	504	4	US-10-342-331-3 Sequence 3, App1
17	385	40.4	1497	4	US-10-157-031-128 Sequence 128, App
18	384	40.3	1466	4	US-10-402-089-4 Sequence 4, App1
19	384	40.3	1466	4	US-10-402-089-6 Sequence 4, App1
20	384	40.3	1466	4	US-10-402-072A-4 Sequence 4, App1
21	384	40.3	1466	4	US-10-402-072A-6 Sequence 6, App1
22	383.5	40.3	1838	5	US-10-852-335A-162 Sequence 162, App
23	383.5	40.3	1838	5	US-10-287-436A-464 Sequence 464, App
24	383.5	40.3	1838	5	US-10-287-436A-1161 Sequence 1161, App
25	382	40.1	1266	4	US-10-367-094-85 Sequence 85, App1
26	382	40.1	1261	4	US-10-367-094-87 Sequence 87, App1
27	382	40.1	1679	4	US-10-367-094-89 Sequence 89, App1

28	381.5	40.1	703	4	US-10-219-449-4 Sequence 4, App1
29	381.5	40.1	703	5	US-10-958-858-4 Sequence 4, App1
30	381.5	40.1	717	4	US-10-219-449-2 Sequence 2, App1
31	381.5	40.1	717	5	US-10-958-858-2 Sequence 2, App1
32	381.5	40.1	733	4	US-10-138-588-40 Sequence 40, App1
33	381.5	40.1	840	4	US-10-367-094-82 Sequence 82, App1
34	380.5	40.0	638	4	US-10-138-588-42 Sequence 42, App1
35	380	39.9	1767	5	US-10-741-600-1476 Sequence 1476, App
36	380	39.9	1767	5	US-10-741-600-1478 Sequence 1478, App
37	380	39.9	1806	5	US-10-741-600-1479 Sequence 1479, App
38	380	39.9	1806	5	US-10-741-600-1475 Sequence 1475, App
39	380	39.9	1818	5	US-10-741-600-1477 Sequence 1477, App
40	378.5	39.8	638	5	US-10-001-887-108 Sequence 108, App
41	378.5	39.8	638	5	US-10-958-863-108 Sequence 108, App
42	378.5	39.8	1745	3	US-09-795-061-4 Sequence 4, App1
43	378.5	39.8	1745	5	US-10-723-860-2660 Sequence 2660, App
44	377.5	39.7	1466	4	US-10-402-089-12 Sequence 12, App1
45	377.5	39.7	1466	4	US-10-402-072A-12 Sequence 12, App1
46	377.5	39.7	1078	4	US-10-058-124-21 Sequence 21, App1
47	377	39.6	1449	4	US-10-402-089-8 Sequence 8, App1
48	377	39.6	1449	4	US-10-402-072A-8 Sequence 8, App1
49	377	39.6	1466	3	US-09-918-715-226 Sequence 226, App
50	377	39.6	1466	3	US-10-177-293-33 Sequence 33, App1
51	377	39.6	1466	4	US-10-177-293-33 Sequence 33, App1
52	377	39.6	1466	4	US-10-301-822-33 Sequence 33, App1
53	377	39.6	1466	4	US-10-257-021-72 Sequence 72, App1
54	377	39.6	1466	4	US-10-357-851-3 Sequence 3, App1
55	377	39.6	1466	4	US-10-358-024-3 Sequence 3, App1
56	377	39.6	1466	4	US-10-734-564-103 Sequence 103, App
57	377	39.6	1466	4	US-10-474-794-226 Sequence 226, App
58	377	39.6	1466	5	US-10-852-335A-159 Sequence 159, App
59	377	39.6	1466	5	US-10-979-159-226 Sequence 226, App
60	377	39.6	1466	5	US-10-287-436A-451 Sequence 451, App
61	377	39.6	1466	5	US-10-287-436A-494 Sequence 494, App
62	377	39.6	1466	5	US-10-287-436A-1151 Sequence 1151, App
63	377	39.6	1466	5	US-10-287-436A-1189 Sequence 1189, App
64	377	39.6	1469	5	US-10-450-763-45550 Sequence 45550, A
65	374	39.3	1518	5	US-10-450-763-53038 Sequence 53038, A
66	373	39.2	544	5	US-10-658-989A-4 Sequence 4, App1
67	373	39.2	1463	4	US-10-402-089-2 Sequence 2, App1
68	372.5	39.1	1463	4	US-10-402-072A-2 Sequence 2, App1
69	372.5	39.1	202	4	US-10-029-386-32935 Sequence 32935, A
70	372.5	39.1	417	5	US-10-096-986-104 Sequence 104, App
71	372.5	39.1	528	5	US-10-719-993-512 Sequence 512, App
72	372.5	39.1	550	5	US-10-719-993-498 Sequence 498, App
73	372.5	39.1	680	4	US-10-295-027-110 Sequence 110, App
74	372.5	39.1	680	5	US-10-643-795A-118 Sequence 118, App
75	372.5	39.1	684	3	US-09-961-403-5 Sequence 5, App1
76	372.5	39.1	684	3	US-10-948-518-118 Sequence 118, App
77	372.5	39.1	693	5	US-10-719-993-502 Sequence 502, App
78	372.5	39.1	720	4	US-10-342-331-4 Sequence 4, App1
79	372.5	39.1	837	4	US-10-096-986-103 Sequence 103, App
80	372	39.1	333	4	US-10-232-175-27 Sequence 27, App1
81	372	39.1	333	6	US-11-139-377-27 Sequence 27, App1
82	372	39.1	333	6	US-11-139-377-27 Sequence 27, App1
83	372	39.1	510	4	US-10-332-175-26 Sequence 26, App1
84	372	39.1	510	6	US-11-139-377-26 Sequence 26, App1
85	372	39.1	660	5	US-10-901-816A-3 Sequence 3, App1
86	372	39.1	660	5	US-10-901-816A-4 Sequence 4, App1
87	372	39.1	662	5	US-10-301-816A-9 Sequence 9, App1
88	372	39.1	662	6	US-10-332-175-33 Sequence 33, App1
89	372	39.1	1014	5	US-11-139-377-33 Sequence 33, App1
90	372	39.1	1014	5	US-10-901-816A-5 Sequence 5, App1
91	372	39.1	1014	5	US-10-901-816A-6 Sequence 6, App1
92	372	39.1	1014	5	US-10-901-816A-7 Sequence 7, App1
93	372	39.1	1014	5	US-10-901-816A-8 Sequence 8, App1
94	372	39.1	1014	5	US-10-901-816A-9 Sequence 9, App1
95	372	39.1	1014	5	US-10-901-816A-10 Sequence 10, App1
96	372	39.1	1014	5	US-10-901-816A-11 Sequence 11, App1
97	372	39.1	1014	5	US-10-901-816A-12 Sequence 12, App1
98	372	39.1	1057	4	US-10-104-889-16 Sequence 16, App1
99	372	39.1	1057	4	US-10-104-889-20 Sequence 20, App1
100	372	39.1	1057	5	US-10-104-793-16 Sequence 16, App1
					Sequence 20, App1

ALIGNMENTS

RESULT 1

US-09-992-600A-58

; Sequence 58, Application US/09992600A

; Publication No. US20030027161A1

; GENERAL INFORMATION:

; APPLICANT: Benjamin, Stephanie

; APPLICANT: Tanaka, Hiroaki

; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF

; FILE REFERENCE: 91.US4.DIV

; CURRENT APPLICATION NUMBER: US/09/992,600A

; PRIOR FILING DATE: 2001-11-13

; PRIOR APPLICATION NUMBER: US 09/924,340

; PRIOR FILING DATE: 2001-08-06

; PRIOR APPLICATION NUMBER: PCT/IB01/01715

; PRIOR FILING DATE: 2001-08-06

; PRIOR APPLICATION NUMBER: US 60/305,456

; PRIOR FILING DATE: 2001-07-13

; PRIOR APPLICATION NUMBER: US 60/302,277

; PRIOR FILING DATE: 2001-06-29

; PRIOR APPLICATION NUMBER: US 60/298,698

; PRIOR FILING DATE: 2001-06-15

; PRIOR APPLICATION NUMBER: US 60/293,574

; PRIOR FILING DATE: 2001-05-25

; NUMBER OF SEQ ID NOS: 114

; SOFTWARE: JPatent

; SEQ ID NO 58

; LENGTH: 163

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-992-600A-58

Query Match 100.0%; Score 952; DB 3; Length 163;

Best Local Similarity 100.0%; Pred. No. 7.4e-54;

Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

; Db 1 MGPPGKGTGHPGLPGPKDCKPGPPGSTRPGABGEPGAMPGQRPGRPPGHVPPGP 60

; Db 1 MGPPGKGTGHPGLPGPKDCKPGPPGSTRPGABGEPGAMPGQRPGRPPGHVPPGP 60

; QY 61 PGQPGAGISAVGLKGDRTGATGERGLAGLPGQPGPPGPPGPGYGMGATGPMGQGGIRG 120

; Db 61 PGQPGAGISAVGLKGDRTGATGERGLAGLPGQPGPPGPPGPGYGMGATGPMGQGGIRG 120

; QY 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; Db 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; QY 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; Db 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; QY 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; Db 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; QY 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; Db 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; QY 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; Db 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; QY 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; Db 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; QY 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; Db 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; QY 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; Db 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; QY 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; Db 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; QY 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; Db 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; QY 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; Db 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; QY 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; SOFTWARE: JPatent

; SEQ ID NO 58

; LENGTH: 163

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-924-340-58

Query Match 100.0%; Score 952; DB 3; Length 163;

Best Local Similarity 100.0%; Pred. No. 7.4e-54;

Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

; Db 1 MGPPGKGTGHPGLPGPKDCKPGPPGSTRPGABGEPGAMPGQRPGRPPGHVPPGP 60

; Db 1 MGPPGKGTGHPGLPGPKDCKPGPPGSTRPGABGEPGAMPGQRPGRPPGHVPPGP 60

; QY 61 PGQPGAGISAVGLKGDRTGATGERGLAGLPGQPGPPGPPGPGYGMGATGPMGQGGIRG 120

; Db 61 PGQPGAGISAVGLKGDRTGATGERGLAGLPGQPGPPGPPGPGYGMGATGPMGQGGIRG 120

; QY 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; Db 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; QY 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; Db 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; QY 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; Db 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; QY 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; Db 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; QY 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; Db 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; QY 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; Db 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; QY 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; Db 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; QY 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; Db 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; QY 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; Db 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; QY 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; Db 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; QY 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; Db 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; QY 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; Db 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; QY 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; Db 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; QY 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; Db 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; QY 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; Db 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; QY 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; Db 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; QY 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; Db 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; QY 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; Db 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; QY 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

; Db 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPPFG 163

US-09-999-570-58
; Sequence 58, Application US/09999570
; Publication No. US20030170628A1
; GENERAL INFORMATION:
; APPLICANT: Benjamin, Stephane
; APPLICANT: Tanaka, Hiroaki
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
; FILE REFERENCE: G-091US08DIY
; CURRENT APPLICATION NUMBER: US/09/999,570
; PRIOR FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 09/924,340
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: PCT/IB01/01715
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/305,456
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/302,277
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/298,698
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/293,574
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: JPatent
; SEQ ID NO 58
; LENGTH: 163
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-999-570-58

Query Match 100.0%; Score 952; DB 3; Length 163;
Best Local Similarity 100.0%; Pred. No. 7,4e-54;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGPPGFKGTGHPGLPGPKDCCGPPPGSTGTPGAEGEPGANGPQGRPGPHVGP 60
DB 1 MGPPGFKGTGHPGLPGPKDCCGPPPGSTGTPGAEGEPGANGPQGRPGPHVGP 60
QY 61 PGQGPAGISAVGLKDDRGATGERSGLAGLPQGPDPGPQGPFGYGMKATGPMGQGI 120
DB 61 PGQGPAGISAVGLKDDRGATGERSGLAGLPQGPDPGPQGPFGYGMKATGPMGQGI 120
QY 121 IPGPBPMGQPGKAGHCNPSDCFGAMPMEQOYPPMTKMGPF 163
DB 121 IPGPBPMGQPGKAGHCNPSDCFGAMPMEQOYPPMTKMGPF 163

RESULT 5
US-10-000-489-58
; Sequence 58, Application US/10000489
; Publication No. US20030092011A1
; GENERAL INFORMATION:
; APPLICANT: Benjamin, Stephane
; APPLICANT: Tanaka, Hiroaki
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
; FILE REFERENCE: 91.US6.DIV
; CURRENT APPLICATION NUMBER: US/10/000,489
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: US 09/924,340
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: PCT/IB01/01715
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/305,456
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/302,277
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/298,698
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/293,574
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: JPatent
; SEQ ID NO 58

; LENGTH: 163
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-000-489-58

Query Match 100.0%; Score 952; DB 4; Length 163;
Best Local Similarity 100.0%; Pred. No. 7,4e-54;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGPPGFKGTGHPGLPGPKDCCGPPPGSTGTPGAEGEPGANGPQGRPGPHVGP 60
DB 1 MGPPGFKGTGHPGLPGPKDCCGPPPGSTGTPGAEGEPGANGPQGRPGPHVGP 60
QY 61 PGQGPAGISAVGLKDDRGATGERSGLAGLPQGPDPGPQGPFGYGMKATGPMGQGI 120
DB 61 PGQGPAGISAVGLKDDRGATGERSGLAGLPQGPDPGPQGPFGYGMKATGPMGQGI 120
QY 121 IPGPBPMGQPGKAGHCNPSDCFGAMPMEQOYPPMTKMGPF 163
DB 121 IPGPBPMGQPGKAGHCNPSDCFGAMPMEQOYPPMTKMGPF 163

RESULT 6
US-10-000-986-58
; Sequence 58, Application US/10000986
; Publication No. US20030096247A1
; GENERAL INFORMATION:
; APPLICANT: Benjamin, Stephane
; APPLICANT: Tanaka, Hiroaki
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
; FILE REFERENCE: 91.US9.DIV
; CURRENT APPLICATION NUMBER: US/10/000,986
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: US 09/924,340
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: PCT/IB01/01715
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/305,456
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/302,277
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/298,698
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/293,574
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: JPatent
; SEQ ID NO 58
; LENGTH: 163
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-000-986-58

Query Match 100.0%; Score 952; DB 4; Length 163;
Best Local Similarity 100.0%; Pred. No. 7,4e-54;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGPPGFKGTGHPGLPGPKDCCGPPPGSTGTPGAEGEPGANGPQGRPGPHVGP 60
DB 1 MGPPGFKGTGHPGLPGPKDCCGPPPGSTGTPGAEGEPGANGPQGRPGPHVGP 60
QY 61 PGQGPAGISAVGLKDDRGATGERSGLAGLPQGPDPGPQGPFGYGMKATGPMGQGI 120
DB 61 PGQGPAGISAVGLKDDRGATGERSGLAGLPQGPDPGPQGPFGYGMKATGPMGQGI 120
QY 121 IPGPBPMGQPGKAGHCNPSDCFGAMPMEQOYPPMTKMGPF 163
DB 121 IPGPBPMGQPGKAGHCNPSDCFGAMPMEQOYPPMTKMGPF 163

RESULT 7
US-10-154-678-58
; Sequence 58, Application US/10154678

```
Publication No. US20030162186A1
GENERAL INFORMATION:
APPLICANT: Benjamin, Stephane
APPLICANT: Tanaka, Hiroaki
TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
FILE REFERENCE: 182.US1.REG
CURRENT APPLICATION NUMBER: US/10/154,678
PRIOR FILING DATE: 2002-10-15
PRIOR APPLICATION NUMBER: US 09/924,340
PRIOR FILING DATE: 2001-08-06
PRIOR APPLICATION NUMBER: US 60/305,456
PRIOR FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: US 60/302,277
PRIOR FILING DATE: 2001-06-29
PRIOR APPLICATION NUMBER: US 60/298,698
PRIOR FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: US 60/293,574
PRIOR FILING DATE: 2001-05-25
NUMBER OF SEQ ID NOS: 112
SOFTWARE: JPatent
SEQ ID NO 58
LENGTH: 163
TYPE: PRT
ORGANISM: Homo sapiens
US-10-154-678-58
```

```
Query Match      100.0%; Score 952; DB 4; Length 163;
Best Local Similarity 100.0%; Pred. No. 7,4e-54;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 MCPPGFKGTGHPGLPGRKDCGKPGPPGSGTRPGAEGEPGAMPGQRPGRGHVGP 60
DB      1 MCPPGFKGTGHPGLPGRKDCGKPGPPGSGTRPGAEGEPGAMPGQRPGRGHVGP 60
QY      61 PQOPGAGISAVGLKGDRCATGREGLAGLPGQPGPPGPGPGPGYGMGATGPMGQGGIFG 120
DB      61 PQOPGAGISAVGLKGDRCATGREGLAGLPGQPGPPGPGPGPGYGMGATGPMGQGGIFG 120
QY      121 IPGPPGPMQPGKAGHCNPSDCFGAMPMEQYPPMKTMKGPRG 163
DB      121 IPGPPGPMQPGKAGHCNPSDCFGAMPMEQYPPMKTMKGPRG 163
```

```
RESULT 8
US-10-838-854-58
Sequence 58, Application US/1083854
Publication No. US20050026182A1
GENERAL INFORMATION:
APPLICANT: Benjamin, Stephane
APPLICANT: Tanaka, Hiroaki
TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
FILE REFERENCE: 91.US6.DIV
CURRENT APPLICATION NUMBER: US/10/838,854
CURRENT FILING DATE: 2004-05-03
PRIOR APPLICATION NUMBER: US/10/000,489
PRIOR FILING DATE: 2001-11-14
PRIOR APPLICATION NUMBER: US 09/924,340
PRIOR FILING DATE: 2001-08-06
PRIOR APPLICATION NUMBER: PCT/IB01/01715
PRIOR FILING DATE: 2001-08-06
PRIOR APPLICATION NUMBER: US 60/305,456
PRIOR FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: US 60/302,277
PRIOR FILING DATE: 2001-06-29
PRIOR APPLICATION NUMBER: US 60/298,698
PRIOR FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: US 60/293,574
PRIOR FILING DATE: 2001-05-25
NUMBER OF SEQ ID NOS: 112
SOFTWARE: JPatent
SEQ ID NO 58
LENGTH: 163
TYPE: PRT
```

```
ORGANISM: Homo sapiens
US-10-838-854-58
```

```
Query Match      100.0%; Score 952; DB 5; Length 163;
Best Local Similarity 100.0%; Pred. No. 7,4e-54;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 MCPPGFKGTGHPGLPGRKDCGKPGPPGSGTRPGAEGEPGAMPGQRPGRGHVGP 60
DB      1 MCPPGFKGTGHPGLPGRKDCGKPGPPGSGTRPGAEGEPGAMPGQRPGRGHVGP 60
QY      61 PQOPGAGISAVGLKGDRCATGREGLAGLPGQPGPPGPGPGPGYGMGATGPMGQGGIFG 120
DB      61 PQOPGAGISAVGLKGDRCATGREGLAGLPGQPGPPGPGPGPGYGMGATGPMGQGGIFG 120
QY      121 IPGPPGPMQPGKAGHCNPSDCFGAMPMEQYPPMKTMKGPRG 163
DB      121 IPGPPGPMQPGKAGHCNPSDCFGAMPMEQYPPMKTMKGPRG 163
```

```
RESULT 9
US-10-723-860-4225
```

```
Sequence 4225, Application US/10723860
Publication No. US20040253606A1
```

```
GENERAL INFORMATION:
APPLICANT: Aziz, Natasha
APPLICANT: Ginsburg, Wendy M.
APPLICANT: Zlotnik, Albert
TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
FILE REFERENCE: 05882.0193.NPUS01
CURRENT APPLICATION NUMBER: US/10/723,860
CURRENT FILING DATE: 2003-11-26
PRIOR APPLICATION NUMBER: 60/429,739
PRIOR FILING DATE: 2002-11-26
NUMBER OF SEQ ID NOS: 8393
SOFTWARE: PatentIn version 3.2
SEQ ID NO 4225
LENGTH: 1603
TYPE: PRT
ORGANISM: Homo sapiens
US-10-723-860-4225
```

```
Query Match      88.3%; Score 840.5; DB 5; Length 1603;
Best Local Similarity 44.5%; Pred. No. 6,5e-46;
Matches 163; Conservative 0; Mismatches 0; Indels 203; Gaps 1;
```

```
QY      1 MCPPGFKGTGHPGLPGRKDCGKPGPPGSGTRPGAEGEPGAMPGQRPGRGHVGP 60
DB      1238 MCPPGFKGTGHPGLPGRKDCGKPGPPGSGTRPGAEGEPGAMPGQRPGRGHVGP 1297
QY      61 PQOPGAGISAVGLKGDRCATGREGLAGLPGQ----- 92
DB      1298 PQOPGAGISAVGLKGDRCATGREGLAGLPGQPGPPGHPGPGEPTDGAAGKEGPPGKQ 1357
QY      93 ----- 92
DB      1358 GFYPPGPKDDBAAGQKQAGEKGRAMPGPGKSGSMGVPPGPPAGERHHPGAPGS 1417
QY      93 ----- 92
DB      1418 GSPGLPGLVPSMGDMVNYDEIRKFRIOELIKMFDERMAVYTSRMQPFMEMAAAPGRPGPP 1477
QY      93 -----GPPRG 97
DB      1478 GNDGAPRGAPGASRGLPQOIGREGQGLPVYRGLPBTGKEKGDIGIGTAGENGGLPGPPG 1537
QY      98 PQOPGAGISAVGLKGDRCATGREGLAGLPGQPGPPGPGPGPGYGMGATGPMGQGGIFG 157
DB      1538 PQOPGAGISAVGLKGDRCATGREGLAGLPGQPGPPGPGPGPGYGMGATGPMGQGGIFG 1597
QY      158 MKGPPG 163
|||||
```

Db 1598 MKGPF 1603

RESULT 10
US-10-797-606-39

; Sequence 39, Application US/10797606
; Publication No. US20050202069A1
; GENERAL INFORMATION:
; APPLICANT: KUROKAWA, Masato
; APPLICANT: NAKAMURA, Hiroaki
; TITLE OF INVENTION: Wound dressing for accelerating epidermal regeneration
; FILE REFERENCE: 292US
; CURRENT APPLICATION NUMBER: US/10/797,606
; CURRENT FILING DATE: 2004-03-11
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 39
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: auxiliary amino acid sequence (Y)
US-10-797-606-39

Query Match 42.5%; Score 405; DB 5; Length 180;
Best Local Similarity 53.4%; Pred. No. 8.4e-19;
Matches 87; Conservative 0; Mismatches 66; Indels 10; Gaps 4;

Qy 2 GPPGFGKGTGHPGLPGPKDCKGRPPGSTGRPCAGBEGPAMGPQGRPPGHVGGPP 61

Db 1 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 60

Qy 62 GPPGPGISAVGKGRGATGERGLAGLPQGRPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 120

Db 61 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 118

Qy 121 IPGPPGPMGPPGKAGHCNPSCDFGAMPMEQQYPPMTMKGPPG 163

Db 119 PPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 154

RESULT 11

US-10-001-843-218
; Sequence 218, Application US/10001843
; Publication No. US20020132255A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Hervé
; APPLICANT: Cafierkey, Robert
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; APPLICANT: Turner, Leah
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and Pr
; FILE REFERENCE: DEX-0267
; CURRENT APPLICATION NUMBER: US/10/001,843
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/249,992
; PRIOR FILING DATE: 2000-11-20
; NUMBER OF SEQ ID NOS: 218
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 218
; LENGTH: 67
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-001-843-218

Query Match 41.5%; Score 395; DB 4; Length 67;
Best Local Similarity 100.0%; Pred. No. 1.6e-18;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 97 GPPGPPGKMGATGPMGQGGIPGIPGPPGPMGQPKAGHCNPSCDFGAMPMEQQYPPMK 156

Db 67 GPPGPPGKMGATGPMGQGGIPGIPGPPGPMGQPKAGHCNPSCDFGAMPMEQQYPPMK 156

Db 1 GPPGPPGKMGATGPMGQGGIPGIPGPPGPMGQPKAGHCNPSCDFGAMPMEQQYPPMK 60

Qy 157 TMKGPPG 163

Db 61 TMKGPPG 67

RESULT 12
US-11-005-609-218

; Sequence 218, Application US/11005609
; Publication No. US20050084899A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Hervé
; APPLICANT: Cafierkey, Robert
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; APPLICANT: Turner, Leah
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and Pr
; FILE REFERENCE: DEX-0267
; CURRENT APPLICATION NUMBER: US/11/005,609
; CURRENT FILING DATE: 2004-12-06
; PRIOR APPLICATION NUMBER: US/10/001,843
; PRIOR FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/249,992
; PRIOR FILING DATE: 2000-11-20
; NUMBER OF SEQ ID NOS: 218
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 218
; LENGTH: 67
; TYPE: PRT
; ORGANISM: Homo sapien
US-11-005-609-218

Query Match 41.5%; Score 395; DB 6; Length 67;
Best Local Similarity 100.0%; Pred. No. 1.6e-18;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 97 GPPGPPGKMGATGPMGQGGIPGIPGPPGPMGQPKAGHCNPSCDFGAMPMEQQYPPMK 156

Db 1 GPPGPPGKMGATGPMGQGGIPGIPGPPGPMGQPKAGHCNPSCDFGAMPMEQQYPPMK 60

Qy 157 TMKGPPG 163

Db 61 TMKGPPG 67

RESULT 13

US-10-723-860-2221
; Sequence 2221, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natsasha
; APPLICANT: Zlorchik, Albert
; APPLICANT: Ginsburg, Wendy M.
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; FILE REFERENCE: 05882.0193.NP0501
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2221
; LENGTH: 910
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-723-860-2221

Query Match 41.0%; Score 390; DB 5; Length 910;
Best Local Similarity 43.2%; Pred. No. 2.9e-17;

Qy 97 GPPGPPGKMGATGPMGQGGIPGIPGPPGPMGQPKAGHCNPSCDFGAMPMEQQYPPMK 156

Db 67 GPPGPPGKMGATGPMGQGGIPGIPGPPGPMGQPKAGHCNPSCDFGAMPMEQQYPPMK 156

Matches 80; Conservative 11; Mismatches 40; Indels 54; Gaps 4;

```
QY 2 GPPGKGTGTHG-----LPGKDDCGKPPPGSTGRGAGEGAMGPOGRPPPGHV 55
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 147 GPPGIPGKMGEGSITMSLPGPKGNPGYPPPGIQLGPTGIPGIPG---PGPPGLM 203
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 56 GPPGPGGPGAPGAGISAVGLKGDGATGERGLA----- 90
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 204 GPPGPGGLPGPKGNMGLNPGPKGEGGLGPPGPPQISEQKRPIDVERQKDDGLP 263
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 91 ---GPPPPGPPG-----YGMGATGPMGQGGIPGTPGPPGMQ 130
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 264 GDRGPPGPIRGPPGPGGEGKEGEGEPKRGKPGKDGEMGQPGIPLGPDGYPGE 323
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 131 PGKAG 135
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 324 PGRDG 328
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

RESULT 14

```
US-10-450-763-35198
; Sequence 35198, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; PRIOR FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 35198
; LENGTH: 1685
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (1159)..(1203)
; OTHER INFORMATION: Synapsins proteins domain identified by eMATRIX, accession
; OTHER INFORMATION: number BL00415N, p-value=4.229e-10, raw score of 4.29
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (42)..(1456)
; OTHER INFORMATION: Collagen triple helix repeat (20 copies) domain identified by
; OTHER INFORMATION: Pfam, accession name Collagen, E-value=3.2e-176, Pfam score of
; OTHER INFORMATION: 598.8
US-10-450-763-35198
```

Query Match 41.0%; Score 390; DB 5; Length 1685;

Best Local Similarity 43.2%; Pred. No. 4.7e-17;

Matches 80; Conservative 11; Mismatches 40; Indels 54; Gaps 4;

```
QY 2 GPPGKGTGTHG-----LPGKDDCGKPPPGSTGRGAGEGAMGPOGRPPPGHV 55
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 147 GPPGIPGKMGEGSITMSLPGPKGNPGYPPPGIQLGPTGIPGIPG---PGPPGLM 203
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 56 GPPGPGGPGAPGAGISAVGLKGDGATGERGLA----- 90
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 204 GPPGPGGLPGPKGNMGLNPGPKGEGGLGPPGPPQISEQKRPIDVERQKDDGLP 263
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 91 ---GPPPPGPPG-----YGMGATGPMGQGGIPGTPGPPGMQ 130
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 264 GDRGPPGPIRGPPGPGGEGKEGEGEPKRGKPGKDGEMGQPGIPLGPDGYPGE 323
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 131 PGKAG 135
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 324 PGRDG 328
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

RESULT 15

```
US-10-450-763-45978
; Sequence 45978, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; PRIOR FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 45978
; LENGTH: 1693
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (1167)..(1211)
; OTHER INFORMATION: Synapsins proteins domain identified by eMATRIX, accession
; OTHER INFORMATION: number BL00415N, p-value=4.229e-10, raw score of 4.29
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (50)..(1462)
; OTHER INFORMATION: Collagen triple helix repeat (20 copies) domain identified by
; OTHER INFORMATION: Pfam, accession name Collagen, E-value=7.1e-185, Pfam score of
; OTHER INFORMATION: 627.6
US-10-450-763-45978
```

Query Match 41.0%; Score 390; DB 5; Length 1693;

Best Local Similarity 43.2%; Pred. No. 4.8e-17;

Matches 80; Conservative 11; Mismatches 40; Indels 54; Gaps 4;

```
QY 2 GPPGKGTGTHG-----LPGKDDCGKPPPGSTGRGAGEGAMGPOGRPPPGHV 55
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 155 GPPGIPGKMGEGSITMSLPGPKGNPGYPPPGIQLGPTGIPGIPG---PGPPGLM 211
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 56 GPPGPGGPGAPGAGISAVGLKGDGATGERGLA----- 90
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 212 GPPGPGGLPGPKGNMGLNPGPKGEGGLGPPGPPQISEQKRPIDVERQKDDGLP 271
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 91 ---GPPPPGPPG-----YGMGATGPMGQGGIPGTPGPPGMQ 130
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 272 GDRGPPGPIRGPPGPGGEGKEGEGEPKRGKPGKDGEMGQPGIPLGPDGYPGE 331
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 131 PGKAG 135
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 332 PGRDG 336
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

Search completed: December 1, 2005, 08:40:59

Job time : 163 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 1, 2005, 08:26:53 ; Search time 11 Seconds
(Without alignments)
70.954 Million cell updates/sec

Title: US-09-924-340-58
Perfect score: 952
Sequence: 1 MGPFGFKGTGHTGFLPGPKG.....GAMPMEQGYPMKTKMGKPG 163

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 26661 seqs, 4788334 residues
Total number of hits satisfying chosen parameters: 26661

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : Published Applications_AA_New:
1: /cgn2_6/prodata/2/pubpaa/US09_NEW_PUB.ppe:*
2: /cgn2_6/prodata/2/pubpaa/US06_NEW_PUB.ppe:*
3: /cgn2_6/prodata/2/pubpaa/US07_NEW_PUB.ppe:*
4: /cgn2_6/prodata/2/pubpaa/US08_NEW_PUB.ppe:*
5: /cgn2_6/prodata/2/pubpaa/PCT_NEW_PUB.ppe:*
6: /cgn2_6/prodata/2/pubpaa/US10_NEW_PUB.ppe:*
7: /cgn2_6/prodata/2/pubpaa/US11_NEW_PUB.ppe:*
8: /cgn2_6/prodata/2/pubpaa/US60_NEW_PUB.ppe:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	385	40.4	1532	6	US-10-821-234-914 Sequence 914, App
2	383.5	40.3	1874	6	US-10-821-234-1182 Sequence 1182, Ap
3	377	39.6	1166	6	US-10-821-234-964 Sequence 964, App
4	372	39.1	1467	6	US-10-821-234-1096 Sequence 1096, Ap
5	356.5	37.4	1366	6	US-10-821-234-1431 Sequence 1431, Ap
6	245	25.7	303	7	US-11-135-855-31 Sequence 31, App1
7	239.5	25.2	406	6	US-10-131-826A-82 Sequence 82, App1
8	238.5	25.1	288	7	US-11-135-855-30 Sequence 30, App1
9	219.5	23.1	253	6	US-10-821-234-1438 Sequence 1438, Ap
10	209	22.0	636	6	US-10-793-626-2922 Sequence 2922, Ap
11	198	20.8	243	6	US-10-131-826A-362 Sequence 362, App
12	165.5	17.4	246	7	US-11-067-121-4 Sequence 4, App1
13	162.5	17.1	119	7	US-11-110-424-4 Sequence 4, App1
14	161.5	17.0	246	7	US-11-067-121-13 Sequence 13, App1
15	161.5	17.0	287	7	US-11-174-150-30 Sequence 30, App1
16	161	16.9	334	6	US-10-514-057-6 Sequence 6, App1
17	157	16.5	724	7	US-11-053-100-41 Sequence 41, App1
18	157	16.5	1174	7	US-11-053-100-43 Sequence 43, App1
19	152.5	16.0	669	7	US-11-053-100-39 Sequence 39, App1
20	149.5	15.7	489	7	US-11-053-100-24 Sequence 24, App1
21	149.5	15.7	504	7	US-11-053-100-28 Sequence 28, App1
22	149.5	15.7	506	7	US-11-053-100-37 Sequence 37, App1
23	149.5	15.7	507	7	US-11-053-100-31 Sequence 31, App1
24	149.5	15.7	507	7	US-11-053-100-34 Sequence 34, App1
25	149.5	15.7	939	7	US-11-053-100-26 Sequence 26, App1

26	149	15.7	879	7	US-11-053-100-55 Sequence 55, App1
27	149	15.7	1225	7	US-11-053-100-50 Sequence 50, App1
28	149	15.7	1339	7	US-11-053-100-56 Sequence 56, App1
29	145	15.2	864	7	US-11-053-100-58 Sequence 58, App1
30	144.5	15.2	624	7	US-11-053-100-48 Sequence 48, App1
31	144.5	15.2	641	6	US-10-848-976-1 Sequence 1, App1
32	144.5	15.2	728	7	US-11-053-100-54 Sequence 54, App1
33	144.5	15.2	735	7	US-11-053-100-44 Sequence 44, App1
34	144.5	15.2	736	7	US-11-053-100-45 Sequence 45, App1
35	144.5	15.2	757	7	US-11-053-100-47 Sequence 47, App1
36	144.5	15.2	774	7	US-11-053-100-49 Sequence 49, App1
37	144.5	15.2	775	7	US-11-053-100-51 Sequence 51, App1
38	144.5	15.2	859	7	US-11-053-100-52 Sequence 52, App1
39	144.5	15.2	879	7	US-11-053-100-57 Sequence 57, App1
40	144.5	15.2	1186	7	US-11-053-100-46 Sequence 46, App1
41	144.5	15.2	1309	7	US-11-053-100-53 Sequence 53, App1
42	144	15.1	288	6	US-10-131-826A-316 Sequence 316, App
43	141.5	14.9	66	6	US-10-514-057-1 Sequence 1, App1
44	141.5	14.9	284	7	US-11-102-240-62 Sequence 62, App
45	141	14.8	334	6	US-10-802-796-728 Sequence 728, App
46	140.5	14.8	574	7	US-11-053-100-40 Sequence 40, App1
47	140.5	14.8	874	7	US-11-053-100-42 Sequence 42, App1
48	136.5	14.3	450	7	US-11-053-100-16 Sequence 16, App1
49	134	14.1	318	6	US-10-802-796-727 Sequence 727, App
50	133	14.0	250	6	US-10-821-234-1237 Sequence 1297, Ap
51	133	14.0	2630	7	US-11-186-731-2 Sequence 2, App1
52	133	14.0	7968	7	US-11-186-731-5 Sequence 36, App1
53	129.5	13.6	356	7	US-11-053-100-36 Sequence 38, App1
54	129.5	13.6	357	7	US-11-053-100-33 Sequence 33, App1
55	129.5	13.6	656	7	US-11-053-100-38 Sequence 35, App1
56	129.5	13.6	657	7	US-11-053-100-35 Sequence 23, App1
57	127.5	13.4	339	7	US-11-053-100-23 Sequence 27, App1
58	127.5	13.4	354	7	US-11-053-100-27 Sequence 25, App1
59	127.5	13.4	357	7	US-11-053-100-30 Sequence 29, App1
60	127.5	13.4	639	7	US-11-053-100-25 Sequence 32, App1
61	127.5	13.4	634	7	US-11-053-100-29 Sequence 1074, Ap
62	127.5	13.4	657	7	US-11-053-100-32 Sequence 955, App
63	126.5	13.3	175	6	US-10-821-234-1074 Sequence 12, App1
64	122	12.8	278	7	US-11-113-424-43 Sequence 43, App1
65	120	12.6	1187	6	US-10-821-234-955 Sequence 57, App1
66	119.5	12.6	100	7	US-11-053-100-17 Sequence 12, App1
67	117.5	12.3	440	7	US-11-102-240-52 Sequence 899, App
68	117	12.3	298	6	US-10-821-234-899 Sequence 28, App1
69	117	12.3	2432	6	US-10-821-234-1240 Sequence 1070, Ap
70	116	12.2	240	7	US-11-021-441-28 Sequence 2, App1
71	116	12.2	1377	6	US-10-821-234-1070 Sequence 2, App1
72	115.5	12.1	963	6	US-10-467-962B-2 Sequence 20, App1
73	114.5	12.0	614	7	US-11-015-546B-20 Sequence 12, App1
74	113	11.9	483	6	US-10-451-375-12 Sequence 372, App
75	111.5	11.7	281	6	US-10-821-234-1032 Sequence 1022, Ap
76	109.5	11.5	552	6	US-10-821-234-1481 Sequence 1481, Ap
77	108	11.3	618	6	US-10-821-234-1481 Sequence 83, App1
78	105	11.0	158	6	US-10-467-962B-83 Sequence 1, App1
79	104.5	11.0	1560	7	US-11-059-982-1 Sequence 79, App1
80	103	10.8	431	7	US-10-821-234-1219 Sequence 1219, Ap
81	102.5	10.8	1133	6	US-10-821-234-444 Sequence 44, App1
82	100	10.5	139	7	US-11-113-424-44 Sequence 881, App
83	99.5	10.5	401	6	US-10-821-234-881 Sequence 1641, Ap
84	99.5	10.5	1970	6	US-10-821-234-1641 Sequence 2, App1
85	99	10.4	417	6	US-10-971-560-2 Sequence 4, App1
86	99	10.4	501	6	US-10-971-560-4 Sequence 10, App1
87	99	10.4	501	7	US-10-971-560-10 Sequence 1, App1
88	99	10.4	501	6	US-11-067-121-1 Sequence 2, App1
89	99	10.4	1618	6	US-10-984-645-2 Sequence 1203, Ap
90	98	10.3	273	6	US-10-821-234-1203 Sequence 1290, App
91	97	10.2	572	6	US-10-821-234-1290 Sequence 158, App
92	95.5	10.0	190	6	US-10-131-826A-158 Sequence 336, App
93	95.5	10.0	283	6	US-10-467-657-336 Sequence 128, App
94	95.5	10.0	575	6	US-10-131-826A-128 Sequence 931, App
95	95	10.0	405	6	US-10-821-234-931 Sequence 1628, Ap
96	94.5	9.9	341	6	US-10-821-234-1628 Sequence 1026, Ap
97	94.5	9.9	406	6	US-10-821-234-1026 Sequence 7, App1
98	94.5	9.9	494	6	US-10-971-560-7 Sequence 7, App1

99 93.5 9.8 524 6 US-10-508-263-60 Sequence 60, Appl
100 93 9.8 280 7 US-11-071-062-3 Sequence 3, Appl

ALIGNMENTS

```
RESULT 1
US-10-821-234-914
; Sequence 914, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; PRIOR FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes Version 1.0
; SEQ ID NO 914
; LENGTH: 1532
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-914

Query Match      40.4%; Score 385; DB 6; Length 1532;
Best Local Similarity 43.3%; Pred. No. 1,7e-23;
Matches 93; Conservative 11; Mismatches 49; Indels 62; Gaps 10;

QY 2 GPPGFKGTGHPGL---PGPKDCKRP-----GPPGSTRPGAGGPPGAMP-----45
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 787 GPPGEGGLGMPRIKRPDPGSDPGKRGLTGPGPGGLPRTPRPGIKGEPAPGKIYVS 846

QY 46 -----GGRPPGHVGPVPPGPPGQPPGAGIS-----AVGLKDRGATGERGLAGLP 90
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 847 EGSSMLTVGPVPPGPPGAMPGRPPGAPGAPGAPGLPGHQEVNLQGPVPPGPPGPPG--P 905

QY 91 GQPPGPPGQPPPGVGYGKMGATGPKGQ-----QGIPIGPVPPG---MGDPGKXGH--136
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 906 SIAPPGPPGPPGPEGLPGPPGPPGSFLSNSSTFLFPPPPGPPGPPGKGDQGPVPPGHQG 965

QY 137 -----CNPSDCFGAMPMEQQYPPMKTKMKGPFG 163
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 966 EQLPGFSTISGSSSFG---LNLQGP-----GPPG 992

RESULT 2
US-10-821-234-1182
; Sequence 1182, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; PRIOR FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes Version 1.0
; SEQ ID NO 1182
; LENGTH: 1874
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1182
```

```
Query Match      40.3%; Score 383.5; DB 6; Length 1874;
Best Local Similarity 39.7%; Pred. No. 2,6e-23;
Matches 83; Conservative 20; Mismatches 55; Indels 51; Gaps 4;

QY 2 GPPGFKGTGHPGLPGPKDCKRPDPGSTRPGABGEPGAMPPOGRPPGPHVGPVPPG 61
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1327 GEPGSGKSGPPPKGERSKESGHSAGAPPPGKPPGDDGPPKSPGVPFPDPPG 1386

QY 62 GQPPGAG-----ISAVGLKDRGAT 81
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1387 GEPGAGQDGPDPKDDGDPQTSPPPTGPPGSPGPPGKXGPPGAPGPEGRQEGKAK 1446

QY 82 GERGLAGLPQPPGPPGPPGPPG-----GMMGATGPMGQGITGIP---GPPGPMQPPGA 134
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1447 GEGAGLEGPPGKGTGPIGPPGAPGKPPGDDGLRGIPGPVGEGLPGSPGPDGPPGMPGPPGLP 1506

QY 135 GHONPDCFGAMPMEQQYPPMKTKMKGPFG 163
   |||:|||||:|||||:|||||:|||||:|||||:
Db 1507 GLKDSGPPKG---EKGHPGLIGLIGPPG 1531
```

```
RESULT 3
US-10-821-234-964
; Sequence 964, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; PRIOR FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes Version 1.0
; SEQ ID NO 964
; LENGTH: 1166
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-964

Query Match      39.6%; Score 377; DB 6; Length 1166;
Best Local Similarity 35.0%; Pred. No. 5,5e-23;
Matches 90; Conservative 8; Mismatches 57; Indels 102; Gaps 6;

QY 2 GPPGFKGTGHPGLPGPKDCKRPDPGSTRPGABGEPGAMPPOGRPPGPHVGP-----57
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 171 GLAGYPGAPGPPGPPGPTSGHPGSPGSGYQGPVGPBGQAQPSGPPPPAIGSPGA 230

QY 58 -----PGPPGQPPGAGI-----SAVGLKDP- 77
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 231 GKDSESGRPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 290

QY 78 -----RGATGERGLAGLP-----GQPPGPPGPPGPPGPPGPPGPPGPPGPPG 104
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 291 GLPGENGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 350

QY 105 -----GRMGATGPMGQGITGIP---GPPGPMQPPGA 146
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 351 GAKGEVGPAGSPGNSGAPGQGEPPGQAGAGQGPVPPGPPGPPGPPGPPGPPGPPGPPG 409

QY 147 PMEQQYPPMKTKMKGPFG 163
   :|||:|||||:|||||:|||||:|||||:|||||:
Db 410 -----PGLMGAGPPG 420

RESULT 4
US-10-821-234-1096
; Sequence 1096, Application US/10821234
```

```
Publication No. US20050255114A1
GENERAL INFORMATION:
APPLICANT: Labat, Ivan
APPLICANT: Stache-Crain, Birgit
APPLICANT: Andarmant, Susan
APPLICANT: Tang, Y. Tom
TITLE OF INVENTION: Methode for Diagnosis and Treatment of Preeclampsia
FILE REFERENCE: 821A
CURRENT APPLICATION NUMBER: US/10/821,234
CURRENT FILING DATE: 2004-04-07
PRIOR APPLICATION NUMBER: US 60/462,047
NUMBER OF SEQ ID NOS: 1704
SOFTWARE: pc_seq_genes Version 1.0
SEQ ID NO 1096
LENGTH: 1467
TYPE: PRT
ORGANISM: Homo sapiens
US-10-821-234-1096

Query Match
Best Local Similarity 39.1%; Score 372; DB 6; Length 1467;
Matches 81; Conservative 7; Mismatches 43; Indels 68; Gaps 6;

QY 2 GPP---GFKGKTGHPGLPGPKGDCGKPPGSTGTPGAESEPGAMGPGQRPGRHVGP 58
DB 872 GPGGATGFPAAARVGPBGSSNAGPPGPPGPAKKEGKGPRTGTGPAGRGVEVGP 931
QY 59 GP-----PGQPPAGIS---AVGLKDRGATGBRGLAIP----- 90
DB 932 GPAGEKSPGADGPAAGTGPFGIAGQGRVGLPQGR---GERGFPGLPGSPGEGKQ 988
QY 91 -----GQRPBPQGPQGPY-----GKMGATGPPGQQ 116
DB 989 GPGSASGEGPPGPPGPPGLAGPPSSGREGAPAAEGSPGRDSSPGAKGDSGTPGAPP 1048
QY 117 GIPGIPGPPGPPMGQPPKAG 135
DB 1049 GARGATGAPGPVPVGPAGKSG 1067

RESULT 5
US-10-821-234-1431
Sequence 1431, Application US/10821234
Publication No. US20050255114A1
GENERAL INFORMATION:
APPLICANT: Labat, Ivan
APPLICANT: Stache-Crain, Birgit
APPLICANT: Andarmant, Susan
APPLICANT: Tang, Y. Tom
TITLE OF INVENTION: Methode for Diagnosis and Treatment of Preeclampsia
FILE REFERENCE: 821A
CURRENT APPLICATION NUMBER: US/10/821,234
CURRENT FILING DATE: 2004-04-07
PRIOR APPLICATION NUMBER: US 60/462,047
PRIOR FILING DATE: 2003-04-07
NUMBER OF SEQ ID NOS: 1704
SOFTWARE: pc_seq_genes Version 1.0
SEQ ID NO 1431
LENGTH: 1366
TYPE: PRT
ORGANISM: Homo sapiens
US-10-821-234-1431

Query Match
Best Local Similarity 37.4%; Score 356.5; DB 6; Length 1366;
Matches 80; Conservative 16; Mismatches 67; Indels 25; Gaps 6;

QY 1 MGPPGKGTGHPGLPGPKDCKGPPGSTGTPGAESEPGAMGPGQRPGRP----- 51
DB 417 MGPPGKGTGHPGLPGPKDCKGPPGSTGTPGAESEPGAMGPGQRPGRP----- 51
QY 52 PGHVGPFGPGGPGAGI-SAVGLKDRGATGBRGLAIPGQGPGRP-----GPPG- 103
```

```
DB 477 PGPIDGARGEBGNIGPFGKPTGDPKNGDKGHAGLAGAGADPGDGNNGAGCPG 536
QY 104 ---YGMKATGPPGQGGIIGIPPPMGQPPGAHCNPSDCG-----AMPMQYPPMK 156
DB 537 QGVQGGKGGPPGPPGFGLPSSGPADEVKPPBGLHGERGLPGLPAGPGRGPPGE 596
QY 157 T-MKGPFG 163
DB 597 SGAAGPTG 604

RESULT 6
US-11-135-855-31
Sequence 31, Application US/11135855
Publication No. US2005025557A1
GENERAL INFORMATION:
APPLICANT: SMITHKLINE BEECHAM CORPORATION
APPLICANT: SMITHKLINE BEECHAM P.L.C.
TITLE OF INVENTION: NOVEL COMPOUNDS
FILE REFERENCE: GP50013
CURRENT APPLICATION NUMBER: US/11/135,855
CURRENT FILING DATE: 2005-05-24
PRIOR APPLICATION NUMBER: US/10/203,708
PRIOR FILING DATE: 2002-08-13
PRIOR APPLICATION NUMBER: PCT/US01/04703
PRIOR FILING DATE: 2001-02-14
PRIOR APPLICATION NUMBER: 60/182,172
PRIOR FILING DATE: 2000-02-14
PRIOR APPLICATION NUMBER: 60/186,084
PRIOR FILING DATE: 2000-02-29
NUMBER OF SEQ ID NOS: 46
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 31
LENGTH: 303
TYPE: PRT
ORGANISM: Homo sapiens
US-11-135-855-31

Query Match
Best Local Similarity 25.7%; Score 245; DB 7; Length 303;
Matches 53; Conservative 10; Mismatches 35; Indels 16; Gaps 4;

QY 13 PGLPGRPGDCKGPPGSTGTPGAESE---PGAMGPPGPPGPPGPPGPPGPPGPPG 69
DB 51 PGLPGR-----PGPPGANGSPGPPHGRITGLPKDGDGKKEGEGKTAGLRKRTPLGL 104
QY 70 SAVGLKDRGATGBRGLAIPGQGPBPQGPBGPPGYGKMGATGPPGQGGIPLGIPG 123
DB 105 A--GKGDQGETTKKPIPEBEGKGEVGPVIGPP-----GPKDGRGEGDPPGLPG 151

RESULT 7
US-10-131-826A-82
Sequence 82, Application US/10131826A
Publication No. US20050245730A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: DeForge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerltsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
```


Query Match 22.0%; Score 209; DB 6; Length 696;
Best Local Similarity 35.6%; Pred. No. 3.6e-10;
Matches 48; Conservative 10; Mismatches 63; Indels 14; Gaps 4;

QY 2 GPPGFGKTKGHPLPBPBKDDCGKPPGSGTGRGAGCEPGAMPQGRPPGPGHVGPPGP 61
DB 520 GPT--PAEPKPKPAEPKPAEPKPAEPKPAEPKPAEPKPAEPKPAEPKPAEPKPAEPK 577
QY 62 GPGG-PAGISAVLKGDRGATGGRGLAGLPQGPDPQGPQGPVGYKMGATGPMGQGIIPG 120
DB 578 AEGGTA-----EPKPAEPKPAEPKPAEPKPAEPKPAEPKPAEPKPAEPKPAEPKPAEP 626
QY 121 IPGPPGMPGQPGKAG 135
DB 627 EPKGPVPEGTPPAQS 641

RESULT 11

US-10-131-826A-362
; Sequence 362, Application US/10131826A
; Publication No. US20050245730A1

GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: DeForge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerltzen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tuma, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin

TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P330R1C128

CURRENT APPLICATION NUMBER: US/10131, 826A

PRIOR FILING DATE: 2002-04-24

PRIOR APPLICATION NUMBER: 60/049911

PRIOR FILING DATE: 1997-06-18

PRIOR APPLICATION NUMBER: 60/056974

PRIOR FILING DATE: 1997-08-26

PRIOR APPLICATION NUMBER: 60/059113

PRIOR FILING DATE: 1997-09-17

PRIOR APPLICATION NUMBER: 60/059115

PRIOR FILING DATE: 1997-09-17

PRIOR APPLICATION NUMBER: 60/059117

PRIOR FILING DATE: 1997-09-17

PRIOR APPLICATION NUMBER: 60/059122

PRIOR FILING DATE: 1997-09-17

PRIOR APPLICATION NUMBER: 60/059184

PRIOR FILING DATE: 1997-09-17

PRIOR APPLICATION NUMBER: 60/059263

PRIOR FILING DATE: 1997-09-18

PRIOR APPLICATION NUMBER: 60/059352

PRIOR FILING DATE: 1997-09-19

PRIOR APPLICATION NUMBER: 60/059588

PRIOR FILING DATE: 1997-09-19

Remaining prior application data removed - See file wrapper or PALM.

NUMBER OF SEQ ID NOS: 550

SEQ ID NO 362

LENGTH: 243

TYPE: PRT

ORGANISM: Homo Sapien

US-10-131-826A-362

Query Match 20.8%; Score 198; DB 6; Length 243;

Best Local Similarity 32.2%; Pred. No. 1.1e-09;
Matches 49; Conservative 6; Mismatches 41; Indels 56; Gaps 4;

QY 11 GHPGLPDKDGGKPPPPSTGRGAGCEPGAMPQGRPPGPGHVGPPGP 70
DB 30 GHPGLPDKDGGKPPPPSTGRGAGCEPGAMPQGRPPGPGHVGPPGP 83
QY 71 AVGLKDRGATGGRGLAGLPQGPDPQGPQGPVGYKMGATGPMGQGIIPG 130
DB 84 -----GAGPAGP-----TGPAEGCSVP----- 101
QY 131 PGKAGHCNPSDFGAMPMEQYPPMKTKMP 162
DB 102 -----PSAFAKRSRVPDPDAPLPF 125

RESULT 12

US-11-067-121-4
; Sequence 4, Application US/11067121
; Publication No. US20050261185A1

GENERAL INFORMATION:

APPLICANT: Martijn, Cecile
APPLICANT: Rondahl, Lena
TITLE OF INVENTION: THERAPEUTIC PROTEINS
FILE REFERENCE: 18909-002001
CURRENT APPLICATION NUMBER: US/11/067,121
PRIOR FILING DATE: 2005-02-25
PRIOR APPLICATION NUMBER: US 60/576,445
PRIOR FILING DATE: 2004-06-02
PRIOR APPLICATION NUMBER: SE 0400489-1
PRIOR FILING DATE: 2004-02-27
NUMBER OF SEQ ID NOS: 20
SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 4

LENGTH: 246

TYPE: PRT

ORGANISM: Mus musculus

US-11-067-121-4

Query Match 17.4%; Score 165.5; DB 7; Length 246;

Best Local Similarity 38.5%; Pred. No. 3.5e-07;

Matches 42; Conservative 7; Mismatches 31; Indels 29; Gaps 5;

QY 14 GLPGRKDGKPPPPSTGRGAGCEPGAMPQGRPPGPGHVGPPGP 73
DB 34 GLPGRKDGKPPPPSTGRGAGCEPGAMPQGRPPGPGHVGPPGP 70

QY 74 LKDRGATGGRGLAGLPQGPDPQGPQGPVGYKMGATGPMGQGIIPG 122
DB 71 LKDRGATGGRGLAGLPQGPDPQGPQGPVGYKMGATGPMGQGIIPG 113

RESULT 13

US-11-110-424-4

; Sequence 4, Application US/11110424

; Publication No. US20050261479A1

GENERAL INFORMATION:

APPLICANT: Hoffmann, Christian K

APPLICANT: Keller, Karsten

TITLE OF INVENTION: A Method for Purifying and Recovering Silk Proteins Using

FILE REFERENCE: C12418 US NA

CURRENT APPLICATION NUMBER: US/11/110,424

PRIOR FILING DATE: 2005-04-20

NUMBER OF SEQ ID NOS: 16

SOFTWARE: PatentIn version 3.2

SEQ ID NO 4

LENGTH: 119

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Monomer of spider silk DP-2A analog protein

US-11-110-424-4

Query Match 17.1%; Score 162.5; DB 7; Length 119;
Best Local Similarity 40.5%; Pred. No. 3.2e-07;
Matches 53; Conservative 7; Mismatches 40; Indels 31; Gaps 10;

QY 4 PGFKKTH-PGLPPKX--DCKKPP--PGSTGRGABEGPAMGPQRPBGHVGPPG 59
DB 15 PQQGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGG 68
QY 60 PPGGPPAGISAVGKGRGATGERGLAGLPQGPBPBGPGGPGGPGGPGGPGGPGG 119
DB 69 GGGGGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGG 110
QY 120 GPGPPGPMGQ 130
DB 111 --PGGYGP-GQ 118

RESULT 14

US-11-067-121-13
; Sequence 13, Application US/11067121
; Publication No. US20050261185A1
; GENERAL INFORMATION:
; APPLICANT: Martijn, Cecile
; APPLICANT: Rondaal, Lena
; TITLE OF INVENTION: THERAPEUTIC PROTEINS
; FILE REFERENCE: 18909-002001
; CURRENT APPLICATION NUMBER: US/11/067,121
; CURRENT FILING DATE: 2005-02-25
; PRIOR APPLICATION NUMBER: US 60/576,445
; PRIOR FILING DATE: 2004-06-02
; PRIOR APPLICATION NUMBER: SE 0400489-1
; PRIOR FILING DATE: 2004-02-27
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 13
; LENGTH: 246
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-067-121-13

Query Match 17.0%; Score 161.5; DB 7; Length 246;
Best Local Similarity 38.6%; Pred. No. 7.1e-07;
Matches 44; Conservative 7; Mismatches 32; Indels 31; Gaps 6;

QY 9 KTGHPGLPGPKGDCGKPPGPGSTGRGABEGPAMGPQRPBGHVGPPGPGGPPAG 68
DB 31 QTC--GLP---PDCSK-----CHGDYVFRGYGPPGP---GPPGIPGNHN-- 70
QY 69 ISAVGLKGRGATGERGLAGLPQGPBPBGPGGPGGPGGPGGPGGPGGPGGPGG 122
DB 71 -----NGNNGATGHEGAKGKGDGLGPR-----GERGQHGPKGKGYGPI 113

RESULT 15

US-11-174-150-30
; Sequence 30, Application US/11174150
; Publication No. US20050260714A1
; GENERAL INFORMATION:
; APPLICANT: Agarwal, Parraj
; APPLICANT: Murooch, Paul R.
; APPLICANT: Rizvi, Safia K.
; APPLICANT: Smith, Randall F.
; APPLICANT: Xiang, Zhaoying
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP50022
; CURRENT APPLICATION NUMBER: US/11/174,150
; CURRENT FILING DATE: 2005-07-01
; PRIOR APPLICATION NUMBER: US/10/257,174
; PRIOR FILING DATE: 2002-10-10
; PRIOR APPLICATION NUMBER: PCT/US01/11797
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 60/196,603

; PRIOR FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: 60/199,417
; PRIOR FILING DATE: 2000-04-24
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO: 30
; LENGTH: 287
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-174-150-30

Query Match 17.0%; Score 161.5; DB 7; Length 287;
Best Local Similarity 36.8%; Pred. No. 8.1e-07;
Matches 43; Conservative 10; Mismatches 29; Indels 35; Gaps 6;

QY 27 PPGST-----GRGABEGPAMGPQRPBGHVGPPGPGGPPAG-----GISA-- 71
DB 17 PPGAHYEMLTGTCRMICDPYSVAPAG--GPAKAP-----PGSTALEVMQDLSNP 69
QY 72 -----VGLKGRGATGERGLAGLPQGPBPBGPGGPGGPGGPGGPGGPGG 123
DB 70 PPPFQGPKGDPGRPKGPPGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGG 118

Search completed: December 1, 2005, 08:38:10
Job time : 12 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 1, 2005, 08:18:41 ; Search time 16 Seconds
(Without alignments)
980.208 Million cell updates/sec

Title: US-09-924-340-58
Perfect score: 952
Sequence: 1 MGPPEKFKGTGHPGLPGPKG.....GAMPMEQGYPPMKTKMGPRG 163

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	840.5	88.3	1603	2 S23810	collagen alpha 1(X
2	390	41.0	1691	1 S22917	collagen alpha 5(I
3	387	40.7	1042	1 CGCH15	collagen alpha 1(I
4	387	40.7	1146	2 A38587	collagen, cornea-s
5	385	40.4	1532	2 A61262	collagen alpha 1(X
6	384	40.3	1549	2 I48103	type VII collagen
7	383	40.2	674	2 S13301	collagen alpha 1(X
8	382.5	40.2	1049	1 CGB075	collagen alpha 1(I
9	382.5	40.2	1838	1 CGH01V	collagen alpha 1(V
10	382.5	40.2	1843	2 S18803	collagen alpha 1(V
11	380.5	40.0	1433	2 A46053	bullous pemphigoid
12	379	39.8	488	2 A27353	collagen alpha 1(I
13	377.5	39.7	674	2 S23297	collagen alpha 1(X
14	377.5	39.7	888	2 S28791	collagen alpha 1(X
15	377.5	39.6	1466	1 CGH07L	collagen alpha 1(I
16	376.5	39.5	1690	1 CGH01B	collagen alpha 4(I
17	376	39.5	886	2 I50694	collagen alpha 1(I
18	374	39.3	921	2 S42617	collagen alpha 1(I
19	374	39.3	1464	2 S59856	collagen alpha 1(I
20	372.5	39.1	680	1 CGH01D	collagen alpha 1(X
21	372.5	39.1	684	2 A53019	collagen alpha 1(X
22	372	39.1	920	2 B34493	collagen alpha 1(I
23	371.5	39.0	730	2 A36226	collagen alpha 1 c
24	371.5	39.0	1806	1 CGH01B	collagen alpha 1(X
25	371	39.0	636	2 S41067	collagen alpha 1(I
26	371	39.0	779	1 CGB015	collagen alpha 1(I
27	371	39.0	1707	2 A33526	collagen alpha 2(I
28	370.5	38.9	1419	2 A41182	collagen alpha 1(I
29	370.5	38.9	1487	2 B41182	collagen alpha 1(I

30	370	38.9	921	2 S40495	collagen alpha 1(I
31	370	38.9	1453	2 S21626	collagen alpha 1(I
32	370	38.9	1464	1 CGH01S	collagen alpha 1(I
33	369.5	38.8	1418	2 T45467	collagen alpha 1(I
34	369.5	38.8	1487	1 CGH06C	collagen alpha 1(I
35	368.5	38.7	2944	2 A54849	collagen alpha 1(V
36	368	38.7	1669	1 CGH04B	collagen alpha 1(I
37	367.5	38.6	1497	2 I49607	procollagen type V
38	367	38.6	1388	2 A53117	collagen alpha 1(X
39	366.5	38.5	615	2 A05269	collagen alpha 1(I
40	366.5	38.5	744	1 A34246	collagen alpha 1(V
41	366.5	38.5	744	1 S23298	collagen alpha 1(V
42	366	38.4	325	2 S02170	collagen alpha 1(I
43	366	38.4	473	2 I50629	collagen - chicken
44	365.5	38.4	453	2 S18804	collagen alpha 4(I
45	365	38.3	680	2 S32436	collagen alpha 2(I
46	365	38.3	680	2 S31216	collagen alpha 1(X
47	365	38.3	754	2 A55267	collagen alpha 5(I
48	364.5	38.3	469	2 A24450	collagen alpha 2(V
49	364.5	38.3	635	2 A57131	collagen alpha 2(V
50	364	38.2	198	2 I49558	collagen alpha 1(I
51	363.5	38.2	633	2 B40983	collagen alpha 1(X
52	363	38.1	920	2 A45748	collagen alpha 1(V
53	363	38.1	1496	1 CGH02V	collagen alpha 2(V
54	362.5	38.1	302	2 T15936	hypothetical prote
55	362.5	38.1	1546	1 CGH02E	collagen alpha 2(X
56	362	38.0	1763	2 S16366	collagen alpha 2(I
57	361.5	38.0	310	2 I50696	collagen alpha 1(I
58	361.5	38.0	677	2 S23296	collagen alpha 2(I
59	361.5	38.0	1024	2 S18251	collagen alpha 1(X
60	361.5	38.0	1142	2 JX0369	collagen alpha 1(X
61	361	37.9	482	2 B31795	collagen alpha 1(X
62	361	37.9	1373	1 A43291	collagen alpha 2(I
63	359.5	37.8	1669	1 CGM54B	collagen alpha 1(I
64	359.5	37.8	931	2 S13580	collagen alpha 1(I
65	359.5	37.8	1414	1 S23809	collagen alpha 2(I
66	359	37.7	671	1 CGR01S	collagen alpha 1(I
67	359	37.7	688	2 A53330	collagen alpha 2(I
68	358	37.6	296	2 A31219	collagen 1 - Caeno
69	357.5	37.6	291	2 T20942	hypothetical prote
70	357.5	37.6	623	2 A45137	collagen alpha 4(I
71	357.5	37.6	743	1 S23779	collagen alpha 1(V
72	357.5	37.6	1670	1 CGH03B	collagen alpha 3(I
73	357	37.5	252	2 A55047	collagen alpha 1(V
74	357	37.5	301	2 T21314	hypothetical prote
75	357	37.5	675	2 S20819	collagen alpha 3(I
76	356.5	37.4	673	1 CGB06C	collagen alpha 1(I
77	356	37.4	1362	2 A40333	collagen alpha 1(I
78	355.5	37.3	1366	1 CGH02S	collagen alpha 2(I
79	355	37.3	290	2 A32249	collagen - sea urc
80	354.5	37.2	295	2 A44984	collagen - nematod
81	354.5	37.2	299	2 T29956	hypothetical prote
82	354	37.2	3124	2 A40020	collagen alpha 1(X
83	353	37.1	1486	1 B40333	collagen alpha 1(I
84	350.5	36.8	85	2 S34665	collagen, cuticula
85	350.5	36.8	964	1 CG428S	collagen alpha 2(I
86	350	36.8	303	2 T28999	hypothetical prote
87	349	36.7	1315	2 A56101	collagen alpha 1(X
88	349	36.7	1774	2 B56101	collagen alpha 1(X
89	348.5	36.6	318	2 S27977	cuticle collagen d
90	348.5	36.6	1758	2 T29350	hypothetical prote
91	348.5	36.6	1759	2 T29351	collagen alpha 2(I
92	348	36.6	1775	2 A31893	collagen alpha 1(I
93	346	36.3	744	2 S15435	collagen alpha 1(V
94	345.5	36.3	516	2 C44479	collagen alpha 1(X
95	345	36.2	742	2 JC7595	scavenger receptor
96	345	36.2	1691	1 CGH06B	collagen alpha 6(I
97	344.5	36.2	298	2 T27644	hypothetical prote
98	344.5	36.2	812	2 S31521	collagen COL1 - f
99	344	36.1	1712	1 CGH02B	collagen alpha 2(I
100	344	36.1	632	2 S42731	collagen alpha 1 c

ALIGNMENTS

RESULT 1
S23810
collagen alpha 1(XVI) chain precursor - human
N:Alternate names: procollagen alpha 1(XVI) chain
C:Species: Homo sapiens (man)
C:Date: 28-Oct-1994 #sequence, revision 28-Oct-1994 #text_change 09-Jul-2004
C:Accession: S23810, PQ0612, S08012
R:Pan, T.C.; Zhang, R.Z.; Mattei, M.G.; Timpl, R.; Chu, M.L.
Proc. Natl. Acad. Sci. U.S.A. 89, 6565-6569, 1992
A:Title: Cloning and chromosomal location of human alpha1(XVI) collagen.
A:Reference number: S23810, MUID:92335339, PMID:1531157
A:Accession: S23810
A:Molecule type: mRNA
A:Residues: 1-1603 <PAN>
A:Cross-references: UNIPROT:007092; UNIPARC:UPI0000126D33; EMBL:M92642; NID:g180757; PID:R:Experimental source: skin fibroblasts
R:Yamaguchi, N.; Kimura, S.; McBride, O.W.; Hori, H.; Yamada, Y.; Kanamori, T.; Yamakoshi, J. Biochem. 112, 856-863, 1992
A:Title: Molecular cloning and partial characterization of a novel collagen chain, alpha A:Reference number: PQ0612, MUID:93203161, PMID:1284248
A:Accession: PQ0612
A:Molecule type: mRNA
A:Residues: 'GGR', 421-536, 'P', 538-1159, 'P', 1161-1162, 'P', 1164, 'P', 1166-1603 <YAN>
A:Cross-references: UNIPARC:UPI000016B3C2; GB:S57132; NID:g298641; PIDN:AAB25797.1; PID:R:Experimental source: placenta
R:Kimura, S.
submitted to the EMBL Data Library, April 1989
A:Description: Partial nucleotide and amino acid sequence of a collagen-like protein from A:Reference number: S08012
A:Accession: S08012
A:Molecule type: mRNA
A:Residues: 403-419, 'GR', 421-536, 'P', 538-846, 'VM', <KIM>
A:Cross-references: UNIPARC:UPI0000073DBB; EMBL:X14963; NID:g29984; PIDN:CAA33085.1; PID:R:Comment: Prolines and lysines at the third position of the tripeptide repeating unit edited and subsequently O-glycosylated.
C:Genetics:
A:Gene: GDB:COL16A1
A:Cross-references: GDB:134045; OMIM:120326
A:Map position: 1p34-1p34
C:Complex: type XVI collagen may be a homotrimer, or a heterotrimer of two alpha 1(XVI) C:Function:
A:Description: structural component of extracellular fibrous polymer as a minor form pro A:Note: may play a role in forming elastic connections at fibril surfaces
C:Keyword: cell binding; coiled coil; extracellular matrix; glycoprotein; hydroxylysine
F:1-21/Domain: signal sequence #status predicted <SIG>
F:122-1603/Product: collagen alpha 1(XVI) chain #status predicted <MAT>
F:22-333/Region: amino-terminal nonhelical #status predicted <NC1>
F:334-1577/Region: interrupted helical
F:334-360/Domain: collagenous COL9 #status predicted <COL9>
F:375-505/Domain: collagenous COL9 #status predicted <COL9>
F:521-554/Domain: collagenous COL8 #status predicted <COL8>
F:539-541/Region: cell attachment (R-G-D) motif
F:572-630/Domain: collagenous COL7 #status predicted <COL7>
F:652-722/Domain: collagenous COL6 #status predicted <COL6>
F:738-875/Domain: collagenous COL5 #status predicted <COL5>
F:887-938/Domain: collagenous COL4 #status predicted <COL4>
F:973-987/Domain: collagenous COL3 #status predicted <COL3>
F:1005-1007/Region: cell attachment (R-G-D) motif
F:1011-1432/Region: collagenous COL2 #status predicted <COL2>
F:1226-1228/Region: cell attachment (R-G-D) motif
F:1472-1577/Domain: collagenous COL1 #status predicted <COL1>
F:1578-1603/Domain: carboxyl-terminal nonhelical #status predicted <NC01>
F:17,327/Binding site: carboxylate (Asn) (covalent) #status predicted

Query Match 88.3%; Score 840.5; DB 2; Length 1603;
Best Local Similarity 44.5%; Pred. No. 4,4e-06;
Matches 163; Conservative 0; Mismatches 0; Indels 203; Gaps 1;

QY 1 MGPFGKGTGHPGLPQPKDCKRPGPGSTGRPGABGEPGAMPGGRGPPGHVPPGP 60

Db 1238 MGPFGKGTGHPGLPQPKDCKRPGPGSTGRPGABGEPGAMPGGRGPPGHVPPGP 1297
QY 61 PQGPAGISAVGLKGDNRGATGERGLAGLPGQ----- 92
Db 1298 PQGPAGISAVGLKGDNRGATGERGLAGLPGQPGPHGPPGPEOTDGAAGEGPPGKQ 1357
QY 93 ----- 92
Db 1358 GFYGPFGKDGAAQKQAGEKGRAGMPGPGKSGMGPVPPGPPAGERGHPGAPGS 1417
QY 93 ----- 92
Db 1418 GSPGLPGVSGMDVNDIEIRFLRQELIKMFDERMAYYTSRMQPMEMAAAPGRGPP 1477
QY 93 -----PGPPG 97
Db 1478 GKDGAPRGAPGSPGLPQOIGREGRQGLPVGRGLPTGKGEKDIGIGIAGENGLPGPPG 1537
QY 98 PQGPFGYGMGATGTMGQGGIRGIPGPPGMPGQPKAGHCNSDPCGAMPMEQYPPMKT 157
Db 1538 PQGPFGYGMGATGHPGQGGIRGIPGPPGMPGQPKAGHCNSDPCGAMPMEQYPPMKT 1597
QY 158 MKGPRG 163
Db 1598 MKGPRG 1603

RESULT 2

S22917

collagen alpha 5(IV) chain precursor, renal splice form - human

N:Alternate names: procollagen alpha 5(IV) chain

N:Contains: collagen alpha 5(IV) chain precursor, leukocyte splice form

C:Species: Homo sapiens (man)

C:Date: 30-Sep-1993 #sequence, revision 27-Feb-1997 #text_change 09-Jul-2004

C:Accession: S22917, A54365; A57079; A37122; I54317; A34850; S18850; I56971; I76598; A35

R:Zhou, J.; Hertz, J.M.; Leinonen, A.; Trygsvaen, K.

J. Biol. Chem. 267, 12475-12481, 1992

A:Title: Complete amino acid sequence of the human alpha-5(IV) collagen chain and identifi

n Alport syndrome patient.

A:Reference number: S22917, MUID:92316923, PMID:1352287

A:Accession: S22917

A:Molecule type: mRNA

A:Residues: 1-967 <ZHO>

A:Cross-references: UNIPROT:P29400; UNIPARC:UPI0000173BDF; GB:M90464; NID:g180826; PIDN:J

R:Zhou, J.; Leinonen, A.; Trygsvaen, K.

J. Biol. Chem. 269, 6608-6614, 1994

A:Title: Structure of the human type IV collagen COL4A5 gene.

A:Reference number: A54365, MUID:94165049, PMID:8120014

A:Accession: A54365

A:Molecule type: DNA

A:Residues: 1-922 <ZH2>

A:Cross-references: UNIPARC:UPI0000173BE0; GB:U04470; NID:9463378; GB:U04520; NID:946342

R:Zhou, J.; Mochizuki, T.; Smeets, H.; Antigmac, C.; Laurila, P.; de Paeye, A.; Trygsvaen

Science 261, 1167-1169, 1993

A:Title: Deletion of the paired alphas5(IV) and alpha5(IV) collagen genes in inherited smc

A:Reference number: A57079, MUID:93561972, PMID:8356449

A:Accession: A57079

A:Molecule type: DNA

A:Residues: 1-27 <ZH4>

A:Cross-references: UNIPARC:UPI00007378A; GB:Z37153; NID:9587203; PIDN:CAA85512.1; PID:G

R:PhilaJantem, T.; Pohjolainen, E.R.; Myers, J.C.

J. Biol. Chem. 265, 13758-13766, 1990

A:Title: Complete primary structure of the triple-helical region and the carboxyl-terminal

A:Reference number: A37122, MUID:90337990, PMID:2380186

A:Accession: A37122

A:Molecule type: mRNA

A:Residues: 84-439, 'GS', 442-624, 'LAIQ', 629-666, 'RR', 669-887, 'R', 889-1264, 1271-1691 <PIH>

A:Cross-references: UNIPARC:UPI0000173BE1; GB:J05558; EMBL:M58526; NID:g1314209

A:Note: submitted to the EMBL Data Library, February 1991

A:Note: the authors translated the codon GCC for residue 115 as Val

R:Remerit, A.; Seri, M.; Myers, J.C.; PhilaJantem, T.; Massella, L.; Rizzoni, G.; De Mai

Hum. Mol. Genet. 1, 127-129, 1992

A:Title: De novo mutation in the COL4A5 gene converting glycine 325 to glutamic acid in f

A:Reference number: 154317, MUID:93244772, PMID:1363780
A:Accession: 154317
A:Status: preliminary, translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 313-324, 'E', 326-330 <REN>
A:Cross-references: UNIPARC:UP1000016B3D0, GB:S59334, NID:g299946, PIDN:AAD13909.1, PID:R.Hochlika, S.L.; Eddy, R.L.; Byers, M.G.; Hoeyhtyee, M.; Shows, T.B.; Trygsvaen, K.
Proc. Natl. Acad. Sci. U.S.A. 87, 1606-1610, 1990
A:Title: Identification of a distinct type IV collagen alpha chain with restricted kidney
A:Reference number: A34850, MUID:90160375, PMID:1689491
A:Accession: A34850
A:Molecule type: mRNA
A:Residues: 914-1264, 1271-1691 <HOS>
A:Cross-references: UNIPARC:UP1000016A70B, EMBL:M31115, NID:g180824, PIDN:AAA52045.1, PID:Genomics 9, 1-9, 1991
A:Title: Characterization of the 3' half of the human type IV collagen alpha-5 gene that
A:Reference number: A37969, MUID:91169491, PMID:2004755
A:Accession: 318850
A:Molecule type: DNA
A:Residues: 924-1264, 1271-1691 <ZH3>
A:Cross-references: UNIPARC:UP1000016A437, EMBL:M63455, EMBL:M63456, EMBL:M63457, EMBL:M63458, EMBL:M63459, EMBL:M63460, EMBL:M63461, EMBL:M63462, EMBL:M63463, EMBL:M63464, EMBL:M63465, EMBL:M63466, EMBL:M63467, EMBL:M63468, EMBL:M63469, EMBL:M63470, EMBL:M63471, EMBL:M63472, EMBL:M63473, NID:g1779292
R.Guo, C.; Van Damme, B.; Van Damme-Lombaerts, R.; Van den Berghe, H.; Cassiman, J.J.; M.
Kidney Int. 44, 1316-1321, 1993
A:Title: Differential splicing of COL4A5 mRNA in kidney and white blood cells: a complex
A:Reference number: 156971, MUID:94133540, PMID:8301933
A:Accession: 156971
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1258-1276 <GUO1>
A:Cross-references: UNIPARC:UP1000016B421, GB:S69168, NID:g545095, PIDN:AAC60612.1, PID:A:Note: kidney splice form
A:Accession: 176598
A:Molecule type: mRNA
A:Status: translated from GB/EMBL/DBJ
A:Residues: 1284-1291, 'T', 'P', 'L', 'G', 'I', 'A', 'C', 'L', 'V', <GUO2>
A:Cross-references: UNIPARC:UP1000011DDFD, GB:S69169, NID:g545097, PIDN:AAC60613.1, PID:R.Meyers, J.C.; Jones, T.A.; Poljansohn, E.R.; Kadri, A.S.; Goddard, A.D.; Sheer, D.; Sch.
Am. J. Hum. Genet. 46, 1024-1033, 1990
A:Title: Molecular cloning of alpha5(V) collagen and assignment of the gene to the region
A:Reference number: A35335, MUID:90252791, PMID:2339699
A:Accession: A35335
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1448-1477 <MYE>
A:Cross-references: UNIPARC:UP10000173BE2
R.Nakazato, H.; Hattori, S.; Ushijima, T.; Matsura, T.; Kotabeahi, Y.; Takada, T.; Yosh.
Kidney Int. 46, 1307-1314, 1994
A:Title: Mutations in the COL4A5 gene in Alport syndrome: a possible mutation in primord
A:Reference number: 156975, MUID:95156893, PMID:7853788
A:Accession: 156975
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1595-1602 <NAK>
A:Cross-references: UNIPARC:UP10000004F8, GB:S75903, NID:g913882, PIDN:AA33374.1, PID:A:Note: premature termination mutation from a patient with Alport syndrome; one other mu
R.Lemliak, H.H.; Schroeder, C.H.; Brunner, H.G.; Nelen, M.R.; Zhou, J.; Trygsvaen, K.;
Genomics 17, 485-489, 1993
A:Title: Identification of four novel mutations in the COL4A5 gene of patients with Alpor
A:Reference number: 154188, MUID:94010948, PMID:8406498
A:Accession: 154188
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1604-1607, 'V', 'H', 'D', 'A', 'Y', 'K', <LEM>
A:Cross-references: UNIPARC:UP100001185C, GB:S65567, NID:g425563, PIDN:AAD13967.1, PID:A:Note: frameshift mutation from a patient with Alport syndrome; five other mutations
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit
ed and subsequently O-glycosylated.
A:Gene: GDB:COL4A5, ATS
A:Cross-references: GDB:120596, OMIM:303630

A:Map position: Xq22-Xq22
A:Introns: 27/3; 47/3; 77/23; 92/3; 107/3; 128/3; 146/3; 155/3; 182/3; 203/3; 215/3; 229/3;
3/3; 799/1; 837/1; 893/1; 923/1; 973/1; 1006/1; 1036/1; 1082/3; 1125/1; 1185/1; 1185/1;
A:Note: The alpha 5(IV) and alpha 6(IV) chain genes are encoded on opposite strands with
C:Complex: This minor type IV collagen is thought to form a heterotrimer of two alpha 5(IV) chains
among trimer amino-terminal domains (with disulfide and desmosine cross-links), dimeric amino
ter associations in the interrupted helical domain (with disulfide and desmosine cross-links)
C:Function:
A:Description: minor structural component of extracellular basement membrane
C:Superfamily: collagen alpha 1(IV) chain
C:Keywords: Alport syndrome; basement membrane; coiled coil; extracellular matrix; glycop
F:1-26/Domain: signal sequence #status predicted <SIG>
F:27-1691/Product: collagen alpha 5(IV) chain, renal splice form #status predicted <MAT1>
F:27-1691/Product: collagen alpha 5(IV) chain, leukocyte splice form #status predicted
F:27-41/Domain: amino-terminal nonhelical, NC2 #status predicted <NC2>
F:42-1462/Region: interrupted helical
F:1463-1691/Domain: carboxyl-terminal nonhelical, NC1 #status predicted <NC1>
F:1473-1573/Domain: collagen IV carboxyl-terminal repeat <CTR1>
F:1583-1687/Domain: collagen IV carboxyl-terminal repeat <CTR2>
F:129-32,38,40,124,451,481,484/Disulfide bonds: interchain #status predicted
F:125/Binding site: carbonyl (Asn) (covalent) #status predicted
F:1482-1570,1515-1573/Disulfide bonds: (or 1482-1573, 1515-1570) #status predicted
F:1527-1533,1638-1644/Disulfide bonds: #status predicted
F:1592-1684,1626-1687/Disulfide bonds: (or 1592-1687, 1626-1684) #status predicted

Query Match 41.0%; Score 390; DB 1; Length 1691;
Best Local Similarity 43.2%; Pred. No. 1,2e-17;
Matches 80; Conservative 11; Mismatches 40; Indels 54; Gaps 4;

Qy 2 GPPGKGTGTHPG-----LPGKGDCCGKPGPGSTGRGABEGPGAMGPGRPGPGHV 55
Db 147 GPPGPGKKGKBEGLTMSLPGPKKNPGYPPGPGIGLPGPTGIPGIPG--GPPGLM 203
Qy 56 GPPGPGPGPAGISAVGLKGRGATGKERGLAGLP-----YGMGATGPMGQGGIPGIPGPGMGQ 130
Db 204 GPPGPGPLGPKGMNGLMFGPKKKGKSGGLGPPPPGQISQKRPIDVFPKGGDGLP 263
Qy 91 --GPPGPPGPGPG-----YGMGATGPMGQGGIPGIPGPGMGQ 130
Db 264 GDRGPPGPPGIRPPGPPGKGEKGEKGEKGEKGEKGRKGRKRGKDXGNGGQIGIPGLPGDPGYPGE 323
Qy 131 PGKAG 135
Db 324 PGRDG 328

RESULT 3
CGCHS
collagen alpha 1(1) chain - chicken (tentative sequence) (fragments)
C:Species: Gallus gallus (chicken)
C:Date: 12-Aug-1981 #sequence revision 06-Jul-1982 #ext_change 31-Mar-2000
C:Accession: A90458; A90181; A02857
R:Highberger, J.H.; Corbett, C.; Dixit, S.N.; Yu, W.; Seyer, J.M.; Kang, A.H.; Gross, J.
Biochemistry 21, 2048-2055, 1982
A:Title: Amino acid sequence of chick skin collagen alpha1(1)-CB8 and the complete primary
A:Reference number: A90458; MUID:82231995; PMID:7093229
A:Accession: A90458
A:Molecule type: protein
A:Residues: 1-1036 <HTG>
A:Cross-references: UNIPARC:UPI0000173B62
A:Experimental source: skin
A:Note: This is the latest in a series of papers from these workers elucidating the sequence
R:Evye, D.R.; Glimcher, M.J.
Biochem. Biophys. Res. Commun. 48, 720-726, 1972
A:Title: Evidence from a previously undetected sequence at the carboxyterminus of the alpha
A:Reference number: A90181; MUID:72243016; PMID:5047697
A:Accession: A90181
A:Molecule type: protein
A:Residues: 1037-1042 <EYR>
A:Cross-references: UNIPARC:UPI0000173B63
A:Experimental source: skin
A:Note: Residues 1037-1042 above correspond to the carboxyl end of the protein
A:Comment: Lysines at positions 103, 700, 934, and 946 above may be hydroxylated in some

C.Comment: Most of the prolines at the third position of the tripeptide repeating unit
C.Comment: Pro-1002 is the only 3-hydroxyproline and the only hydroxylated proline in pc
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C:Keywords: coiled coil; extracellular matrix; glycoprotein; pyroglutamic acid; trimer;
F1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

Query Match 40.7%; Score 387; DB 1; Length 1042;
Best Local Similarity 46.9%; Pred. No. 1.3e-17;
Matches 82; Conservative 12; Mismatches 59; Indels 22; Gaps 5;

QY 2 GPP---GFGKTKGHPGLPGPKDCKGKPGPGSTGRPGAGBPGAMQPGRPQPGHVGPP 58
DB 707 GPPGATGFGAAGKRVPPSPSGNIGLPBPFGPAKGSKPGGTGPAPRPGEPGAGPP 766
QY 59 GPPGPGPAG---ISAVGLKDRGATGEGLAGLPQ-----PQPPGPGPGYGYKMG 108
DB 767 GPPGEGKSGADGDPICAPGTPGQGIAGQGVGLPGQGERGFPGLPQPSGPP--GKQG 824
QY 109 ATGPMQGGTIPGTPGPGMGQPGKAGHCNPSGCFAMMEQOYPPMTKMPFG 163
DB 825 PGGASGERPQPMGPPGLAGPPGAGRGAPGAGSA-----PGRDGAAGPKG 872

RESULT 4

A38587
collagen, cornea-specific - chicken

C.Species: Gallus gallus (chicken)

C.Date: 14-Feb-1992 #sequence_revision 15-Aug-1997 #text_change 09-Jul-2004

C.Accession: S16501; A38587

R.Marchant, J.K.; Linsenmayer, T.F.; Gordon, M.K.

Proc. Natl. Acad. Sci. U.S.A. 88, 1560-1564, 1991

A>Title: cDNA analysis predicts a cornea-specific collagen.

A.Reference number: A38587; MUID:91142213; PMID:1705041

A.Accession: S16501

A.Molecule type: mRNA

A.Residues: 1-1146 <MAR>

A.Cross-references: UNIPROT:Q90584; UNIPARC:UPI000006PC061; EMBL:ME0172; NID:g211609; PID

A.Accession: A38587

A.Molecule type: mRNA

A.Residues: 1-174,'X',176-233,'X',235-344,'X',346-408,'X',410-499,'X',501-876,'X',878-11

A.Cross-references: UNIPARC:UPI000017A187; GB:ME0172

C:Keywords: cornea

Query Match 40.7%; Score 387; DB 2; Length 1146;
Best Local Similarity 40.1%; Pred. No. 1.4e-17;
Matches 89; Conservative 10; Mismatches 57; Indels 66; Gaps 7;

QY 2 GPPGFGKTKGHPGLPGPKDCKGKPGPGSTGRPGAGBPGAMGP-----QGRP 49
DB 350 GPPGPGSDTGEPLTGPQ--GPPGLPAGNPRGAPGAPGAPKVISAEGSTTIALPGRP 406
QY 50 GPPGHVPPPGPQPGPAGISAVGLKDRGATGERG-----LAG 88
DB 407 GPPGPGIPGPGVPBPV--PAGLPQGQSPRGKSNABVETIKTEVSSLASGMSLD 464
QY 89 LFGQPPPPPPQPGPGYCKMGATGPMGQGP-----GIRPPPP 127
DB 465 LQGRAGPPPPPPGPGSSVGLTGPFRPGLPSPGPGPGSGSVSTSTFTVSGPPPPPP 524
QY 128 MQQPGKAGHCNPSDCCGAMPMEQOYPPMK-----TMKPPFG 163
DB 525 PPGKDGEGPFRG-FTGEPGEPLDGFSSHGCTVTMQGPPG 565

RESULT 5

A61262
collagen alpha 1(XVII) chain - human (fragment)

N/Alternate names: bullous pemphigoid 180K autoantigen BPAG2; bullous pemphigoid antigen
C:Species: Homo sapiens (man)

C.Date: 12-May-1994 #sequence_revision 12-Jul-1996 #text_change 09-Jul-2004

C.Accession: I56325; I55345; A61262

R.Giudice, G.J.; Emery, D.J.; Diaz, L.A.

J. Invest. Dermatol. 99, 243-250, 1992

A>Title: Cloning and primary structural analysis of the Bullous pemphigoid autoantigen, I

A.Reference number: I56325; MUID:92281323; PMID:1324962

A.Accession: I56325

A>Status: preliminary; translated from GB/EMBL/DBJ

A.Molecule type: mRNA

A.Residues: 1-1532 <RES>

A.Cross-references: UNIPROT:Q9JMD9; UNIPARC:UPI000006F673; GB:M91669; NID:g179516; PIDN:

R.J.L.; K.H.; Sawamura, D.; Giudice, G.J.; Diaz, L.A.; Matzel, M.G.; Chu, M.L.; Utico, J.

J. Biol. Chem. 266, 24064-24069, 1991

A>Title: Genomic organization of collagenous domains and chromosomal assignment of human

A.Reference number: I55345; MUID:92084712; PMID:1746679

A.Accession: I55345

A>Status: preliminary; translated from GB/EMBL/DBJ

A.Molecule type: mRNA

A.Residues: 543-890,'P' <RES>

A.Cross-references: UNIPARC:UPI000006F72A; GB:M63730; NID:g179520; PIDN:AA51839.1; PID:

R.Giudice, G.J.; Squitiera, H.L.; Elias, P.M.; Diaz, L.A.

J. Clin. Invest. 87, 734-738, 1991

A>Title: Identification of two collagen domains within the bullous pemphigoid autoantigen

A.Reference number: A61262; MUID:91123476; PMID:1846881

A.Accession: A61262

A.Molecule type: mRNA

A.Residues: 543-890,'P' <GIU>

A.Cross-references: UNIPARC:UPI000006F72A

C:Genetics:

A:Gene: GDB:COL17A1; BPAG2; BP180

A:Cross-references: GDB:131396; OMIM:113811

A.Map position: 10q24.3-10q24.3

Query Match 40.4%; Score 385; DB 2; Length 1532;
Best Local Similarity 43.3%; Pred. No. 2.4e-17;
Matches 93; Conservative 11; Mismatches 49; Indels 62; Gaps 10;

QY 2 GPPGFGKTKGHPGL--PQPKDCKGP-----GPPGSGRPGAGBPGAMGP---- 45
DB 787 GPPGQGLTGMPIGIRPGPSGDPKPGGLTGPQGGP.GTGRPRGIGEPGAPKIVTS 846
QY 46 -----QGRPGPGHVGPPPPQGPAGIS-----AVGLKDRGATGERGAGLP 90
DB 847 EGSSMLTVPGPDPGAMGPPGPAGPAGAGLPQHGVNLTQGPDPGPPGRGPPG-P 905
QY 91 GQPPGPPGPGPPGPKMGATGPMGO-----QGIPGIRPPGP--MGQPGKAGH-- 136
DB 906 SIQPPGPPGPPGEGLPBPBPSSFLNSSTFTLGPBPBPBPBPBPBPBPBPBPBP 965
QY 137 -----CNPSDCGAMPMEQOYPPMTKMPFG 163
DB 966 EQGLPGFTSSSSSFG--LNLQGRP-----GPPG 992

RESULT 6

I48103
type VII collagen - Chinese hamster (fragment)

C:Species: Cricetus griseus (Chinese hamster)

C.Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 09-Jul-2004

C.Accession: I48103

R.Greenspan, D.S.

Hum. Mol. Genet. 2, 273-278, 1993

A>Title: The carboxyl-terminal half of type VII collagen, including the non-collagenous

A.Reference number: I48103; MUID:93271985; PMID:8499916

A.Accession: I48103

A>Status: preliminary; translated from GB/EMBL/DBJ

A.Molecule type: mRNA

A.Residues: 1-1549 <RES>

A.Cross-references: UNIPROT:Q60444; UNIPARC:UPI00000E753D; GB:L06863; NID:g386624; PIDN:

F.1484-1536/Domain: animal knitr-type proteinase inhibitor homology <BPI>

Query Match 40.3%; Score 384; DB 2; Length 1549;
Best Local Similarity 45.9%; Pred. No. 2.8e-17;
Matches 78; Conservative 13; Mismatches 49; Indels 30; Gaps 4;

QY 2 GPPGFGKTKGHPGLPGPKDCKGKPGPGSTGRPGAGBPGAMGPPGPPHVGPPPP 61
DB 100 GPPGFGKTKGHPGLPGPKDCKGKPGPGSTGRPGAGBPGAMGPPGPPHVGPPPP 61

QY 56 GPPGPPGPGPAGI-----SAVGLKGD-----R 78
Db 87 GPPGPPGPKGPGAGKRGFGDGRNGEKGPPGAGLKGKNGVGDGEDGAPGPGMR 146
QY 79 GATGRRGAGLP-----GQPGPPGPGPPGY-----104
Db 147 GATGRRGRRGLPGAAGARGNDARSGSDQPGPPGPGTAAFGGSGPGAKGVGPAGSPSS 206
QY 105 ---GKMGATGPMGQGGIPIGPPGPPMGQPGKAGHCNPSDCTFGAMPMEQGYPMKTKMGP 161
Db 207 GATGRRGRRGLPGAAGARGNDARSGSDQPGPPGPGTAAFGGSGPGAKGVGPAGSPSS 259
QY 162 PG 163
Db 260 PG 261

RESULT 9
CGHUV
collagen alpha 1(V) chain precursor - human
N:Alternate names: procollagen alpha 1(V) chain
C:Species: Homo sapiens (man)
C:Date: 22-Nov-1993 #sequence revision 03-Oct-1995 #text change 09-Jul-2004
C:Accession: S18802; A61142; S11303; S03978; S43642; S58665
R:Greenspan, D.S.; Cheng, W.; Hofman, G.G.
J. Biol. Chem. 266, 24727-24733, 1991
A:Title: The pro-alpha1(V) collagen chain. Complete primary structure, distribution of e
A:Reference number: S18802; MUID:92105142; PMID:1722213
A:Accession: S18802
A:Molecule type: mRNA
A:Residues: 1-1838 <GRE>
A:Cross-references: UNIPARC:UPI000006E992; GB:M76729; NID:G189519; PIDN:
R:Takahara, K.; Sato, Y.; Okazawa, K.; Okamoto, N.; Noda, A.; Yaot, Y.; Kato, I.
J. Biol. Chem. 266, 13124-13129, 1991
A:Title: Complete primary structure of human collagen alpha-1(V) chain.
A:Reference number: S16024; MUID:91302336; PMID:2071595
A:Accession: S16024
A:Molecule type: mRNA
A:Residues: 1-81, 'Q', '84-389, 'A', '391-676, 'K', '678-1294, 'PS', '1297, 'RS', '1300-1553, 'R', '1555-
A:Cross-references: UNIPARC:UPI0000126DID; GB:D90279; NID:G219509; PIDN:BAAL4323.1; PID:
A:Note: parts of this sequence were determined by protein sequencing
R:Yaot, Y.; Hashimoto, K.; Takahara, K.; Kato, I.
Exp. Cell Res. 194, 180-185, 1991
A:Title: Insulin binds to type V collagen with retention of mitogenic activity.
A:Reference number: A61142; MUID:91224163; PMID:1709100
A:Accession: A61142
A:Molecule type: protein
A:Residues: 823-824, 'X', '826-842 <YAO>
A:Cross-references: UNIPARC:UPI0000173BB7
A:Note: the residue designated 'X' is probably glycosylated hydroxylysine; this cyanogen
R:Yaot, Y.; Hashimoto, K.; Koitabashi, H.; Takahara, K.; Ito, M.; Kato, I.
Biochem. Biophys. Acta 1035, 139-145, 1990
A:Title: Primary structure of the heparin-binding site of type V collagen.
A:Reference number: S11303; MUID:90366601; PMID:2203476
A:Accession: S11303
A:Molecule type: protein
A:Residues: 823-824, 'X', '826-848, 'I', '850-851, 'P', '853, 'PR', '856-893, 'D', '895-932, 'X', '934-950
A:Cross-references: UNIPARC:UPI0000173BB8
A:Note: the residues designated 'X' are probably glycosylated hydroxylysine; this sequen
R:Seyer, J.M.; Kang, A.H.
Arch. Biochem. Biophys. 271, 120-129, 1989
A:Title: Covalent structure of the collagen amino acid sequence of three cyanogen bromide-d
A:Reference number: S03978; MUID:89227189; PMID:2496661
A:Accession: S03978
A:Molecule type: protein
A:Residues: 621-640, 'G', '642-649, 'L', '651-662, 'E', '664-667, 'Q', '669-676, 'Q', '678-683, 'P', '685-
<SEY>
A:Cross-references: UNIPARC:UPI0000173BB9
A:Note: there are a number of inconsistencies in the sequences in figures 6 and 7,
R:Morzidi-Ameli, M.; Rousseau, J.C.; Klemm, J.P.; Champilaud, M.F.; Bouillon, M.M.; Ben
Eur. J. Biochem. 221, 987-995, 1994
A:Title: Diversity in the processing events at the N-terminus of type-V collagen.

A:Reference number: S43642; MUID:94237164; PMID:8181482
A:Accession: S43642
A:Molecule type: protein
A:Residues: 565-576, '576-758, 'X', '760-763, 'X', '765-772, '1012-1029, '1219-1232, '1465-1474, 'X', '14'
A:Cross-references: UNIPARC:UPI0000173BB4; UNIPARC:UPI0000173BBB; UNIPARC:UPI0000173BBC;
R:Feessler, L.I.; Brosh, S.; Chapin, S.; Feessler, J.H.
J. Biol. Chem. 261, 5034-5040, 1986
A:Title: Tyrosine sulfation in precursors of collagen V.
A:Reference number: A56977; MUID:86168226; PMID:3082875
A:Contents: annotation; identification of tyrosine sulfate in the amino-terminal propept
R:Lee, S.; Greenspan, D.S.
Biochem. J. 310, 15-22, 1995
A:Title: Transcriptional promoter of the human alpha-1(V) collagen gene (COL5A1).
A:Reference number: S58665; MUID:95374437; PMID:7646438
A:Accession: S58665
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-36 <LEB>
A:Cross-references: UNIPARC:UPI000007363D; GB:L38808; NID:G1020325; PIDN:AAA79853.1; PID:
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit (C
are 5-hydroxylated and subsequently O-glycosylated.
C:Comment: A long form of the mature protein containing part of the amino-terminal proper
ile the heterotrimers are probably processed to the long form.
C:Genetics:
A:Gene: GDB:COL5A1
A:Cross-references: GDB:131457; OMIM:120215
A:Map position: 9q34.2-9q34.3
C:Complex: type V collagen may be a homotrimer of alpha 1(V) chains, a heterotrimer of ty
alpha 2(V) chain and one alpha 3(V) chain, initially linked by disulfide bonds among the
length, is formed with desmosome cross-links made from lysine and allysine residues
C:Function:
A:Description: structural component of extracellular fibrous polymer associated with cell
A:Note: may play a role in controlling the lateral growth of collagen I fibrils
C:Superfamily: collagen alpha 1(V) chain; fibrillar collagen carboxyl-terminal homology
C:Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyprolin
F:1-37/Domain: signal sequence #status predicted <SIG>
F:36-261/Domain: PARP-like #status predicted <PARP>
F:38-541/Domain: amino-terminal propeptide #status predicted <PRO>
F:542-1605/Product: collagen alpha 1(V) chain, short form #status predicted <MAT>
F:542-588/Region: amino-terminal nonhelical telopeptide
F:559-1572/Region: helical
F:645-647/Region: cell attachment (R-G-D) motif
F:663-665/Region: cell attachment (R-G-D) motif
F:897-929/Region: heparin binding
F:1573-1605/Region: carboxyl-terminal nonhelical telopeptide
F:1606-1838/Domain: carboxyl-terminal propeptide #status predicted <CPR>
F:1615-1837/Domain: fibrillar collagen carboxyl-terminal homology <FCC>
F:38/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted
F:62-244, 183-237/Disulfide bonds: #status predicted
F:159, 176, 385, 1672, 1741/Binding site: carboxylate (Asn) (covalent) #status predicted
F:234, 236, 240, 262, 263, 273, 274, 275, 277, 279, 280, 338, 340, 346, 347, 352, 357, 416, 417, 420, 421/Bit
F:535/Modified site: allysine (Lys) #status predicted
F:541-542/Cleavage site: Ala-Gln (procollagen N-endopeptidase) #status predicted
F:542/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted
F:570, 576, 621, 639, 648, 654, 657, 675, 678, 690, 693, 696, 705, 717, 720, 726, 732, 741, 750, 753, 756, 76;
site: 4-hydroxyproline (Pro) #status experimental
F:627, 642, 687, 708, 744, 774, 795, 804, 807, 810, 819, 825, 846, 864, 882, 897, 1482/Binding site: 5-hydro
F:627, 642, 687, 774, 795, 804, 807, 810, 819, 825, 846, 864, 882, 897, 1482/Binding site: carboxylate
F:708, 744/Binding site: carboxylate (Lys) (covalent) #status experimental
F:1482/Modified site: 5-hydroxylysine (Lys) #status predicted
F:1605-1606/Cleavage site: Ala-Asp (procollagen C-endopeptidase) #status predicted
F:1639, 1645, 1662, 1671/Disulfide bonds: interchain #status predicted
F:1680-1835, 1746-1789/Disulfide bonds: #status predicted

Query Match 40.2%; Score 382.5; DB 1; Length 1838;
Best Local Similarity 39.7%; Pred. No. 4e-11;
Matches 83; Conservative 19; Mismatches 56; Indels 51; Gaps 4;

QY 2 GPPGRRGRRGLPGAAGARGNDARSGSDQPGPPGPGTAAFGGSGPGAKGVGPAGSPSS 61
Db 1291 GPPGRRGRRGLPGAAGARGNDARSGSDQPGPPGPGTAAFGGSGPGAKGVGPAGSPSS 1350
QY 62 GPPGRRGRRGLPGAAGARGNDARSGSDQPGPPGPGTAAFGGSGPGAKGVGPAGSPSS 81

C>Date: 07-Oct-1994 #sequence revision 10-Nov-1995 #text_change 09-Jul-2004
 C/Accession: S23297; A31896; S65594; S77711; I50218
 R:Nimomlya, Y.; Castagnola, P.; Gerecke, D.; Gordon, M.K.; Jacenko, O.; LuValle, P.; McDaniel, N.; Olsen, B.R.
 A/Title: The molecular biology of collagens with short triple-helical domains.
 A/Reference number: S22243
 A/Accession: S23297
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-674 <NIN>
 A/Cross-references: UNIPROT:P08125; UNIPARC:UPI00001773E7
 R:LuValle, P.; Nimomlya, Y.; Rosenblum, N.D.; Olsen, B.R.
 J. Biol. Chem. 263, 18378-18385, 1988
 A/Title: The type X collagen gene. Intron sequences split the 5'-untranslated region and
 A/Reference number: A31896; MUID:89054019; PMID:2461368
 A/Accession: A31896

A/Molecule type: mRNA
 A/Residues: 1-75 <LUV>
 A/Cross-references: UNIPARC:UPI00001773E8
 R:Nimomlya, Y.; Gordon, M.; van der Reest, M.; Schmidt, T.; Linsemayer, T.; Olsen, B.R.
 J. Biol. Chem. 261, 5041-5050, 1986
 A/Title: The developmentally regulated type X collagen gene contains a long open reading
 A/Reference number: I50218; MUID:86168227; PMID:3082876
 A/Accession: S65594
 A/Molecule type: DNA
 A/Residues: 'T','9','D','11-12','EDQMKLTILEFM','30-31','TCKSGRAFTTYMILQNMADLVSSHT','48-89','L','
 629','POAVLSLIMWTICGSCIONMWSIPLMFLISQVSYLSNNIPLTMS' <NIN1>
 A/Cross-references: UNIPARC:UPI00001773E9; EMBL:M13496; NID:g211699; PIDD:AAA48736.1; PI
 A/Accession: S77711
 A/Molecule type: Protein
 A/Residues: 104-112,'X',114-117,453-466 <NIN2>
 A/Cross-references: UNIPARC:UPI00001773EA; UNIPARC:UPI00001773EB
 C:Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology
 C/Keywords: coiled coil; extracellular matrix; glycoprotein; homotrimer; hydroxyproline;
 F:1-18/Domain: signal sequence #status predicted <SIG>
 F:547-673/Domain: complement C1q carboxyl-terminal homology <CIQ>
 F:453,456/Modified site: hydroxyproline (Pro) #status experimental
 F:611/Binding site: carbohydrate (Aan) (covalent) #status predicted

Query Match 39.7%; Score 377.5; DB 2; Length 674;
 Best Local Similarity 40.0%; Pred. No. 3.6e-17;
 Matches 90; Conservative 9; Mismatches 55; Indels 71; Gaps 7;

QY 2 GPPGFK-GKTGHPGLP-----GPKGDCGKGGPGST-GR 33
 DB 91 GPPGSTVGRKQMPGLPKRGKNGEKGEAGVGLPGARGGPPGI.PGPAGLSVPGK 150
 QY 34 PBAEGEPGAMGPPGRRGPPGH-----VGPPGPPGPPGA 67
 DB 151 PGPQGPFGAOGPPRGKRGKBERPVGINGOKGEMGFVGRPGNRLPGPQGGULPSPA 210
 QY 68 GI-----SAVGLKGRGATGERGLAGLPQPQPPGPPGPGYKGMATGPMQCGI 118
 DB 211 GIGKPGENGLPGOPGKGRGLPGARAGEGICIPGQPPPSPEVGIQKGPMPGPPPAIG 270
 QY 119 PGIPGPPGPMGPPGKAGHCNPSPDCFGAMPMEQOYPPMKTKMGPPG 163
 DB 271 PBAKGLPGPAGLPGSPG-----LPFGF-----KPLPGMKKGHRGPEG 307

RESULT 14

S28791
 collagen alpha 1(XI) chain - chicken (fragment)
 C/Species: Gallus gallus (chicken)
 C/Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 09-Jul-2004
 C/Accession: S28791
 R:Nah, H.D.; Barembaum, M.; Upholt, W.B.
 J. Biol. Chem. 267, 22581-22586, 1992
 A/Title: The chicken alpha(XI) collagen gene is widely expressed in embryonic tissues.
 A/Reference number: S28791; MUID:93054557; PMID:1429667
 A/Accession: S28791
 A/Status: preliminary

A/Molecule type: mRNA
 A/Residues: 1-888 <NNA>
 A/Cross-references: UNIPROT:Q90796; UNIPARC:UPI00000FBDJ3; EMBL:M88593; NID:g211619; PIDD
 C:Superfamily: collagen alpha 1(V) chain; fibrillar collagen carboxyl-terminal homology
 F:665-887/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 39.7%; Score 377.5; DB 2; Length 888;
 Best Local Similarity 47.3%; Pred. No. 4.6e-17;
 Matches 80; Conservative 14; Mismatches 56; Indels 19; Gaps 4;
 QY 2 GPPGFKTGHPLGPPKDCGKDPGPGSTGRPGAEGEPGAMGPPGPPGPPGPP 61
 DB 343 GNPGRSGSGTSGPGRGKGEAGPPGAAGPPGAKGPPGDDGKGNPVPVFPDPPPP 402
 QY 62 GQPPAGISAVGLKGRATGERGLAGLPQGPSPGPPQ-----GPPG-YGKMGATGPMG 114
 DB 403 GEPGAGDGVG-----GEKEDGDPGQPPGPPGEGAGPPGPPGKRRPPGATGAE 454
 QY 115 QGGIPGIPGPPGPMGQPKAGHCNPSPDCFGAMPMEQOYPPMKTKMGPPG 163
 DB 455 RQEGKAGKGEAGAPKTPVGRGPAKRGPE---GLRGIPGPPV 499

RESULT 15

collagen alpha 1(III) chain precursor - human
 N/Alternate names: procollagen alpha 1(III) chain
 C/Species: Homo sapiens (man)
 C/Date: 24-Apr-1984 #sequence_revision 01-Sep-1995 #text_change 09-Jul-2004
 C/Accession: S05272; S04642; PE0011; S01726; S04887; A90399; A94562; I51868; S59511; A904
 R:Prockop, D.J.
 Submitted to the EMBL Data Library, February 1989
 A/Reference number: S05272
 A/Accession: S05272
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-1240,'V',1242-1466 <PRC>
 A/Cross-references: UNIPROT:P02461; UNIPARC:UPI0000000CDE; EMBL:X14420; NID:g30057; PIDD:
 R:Ala-Koko, L.; Kontusaari, S.; Baldwin, C.T.; Kuivaniemi, H.; Prockop, D.J.
 Biochem. J. 260, 509-516, 1989
 A/Title: Structure of cDNA clones coding for the entire prepro-alpha1(III) chain of human
 A/Accession: S04642; MUID:89350838; PMID:2764886

A/Accession: S04642
 A/Molecule type: mRNA
 A/Residues: 1-1196 <ALA>
 A/Cross-references: UNIPARC:UPI0000173880; EMBL:X14420; NID:g30057; PIDD:CAA32583.1; PID:
 A/Note: the complete sequence is not shown
 R:Benson-Chanda, V.; Su, M.W.; Weil, D.; Chu, M.L.; Ramirez, F.
 Gene 78, 255-265, 1989
 A/Title: Cloning and analysis of the 5' portion of the human type-III procollagen gene (C
 A/Reference number: PE0011; MUID:89378752; PMID:2777083
 A/Accession: PE0011
 A/Molecule type: DNA
 A/Residues: 1-176 <BEN>
 A/Cross-references: UNIPARC:UPI000016A703; GB:M26939; NID:g180813; PIDD:AAA52040.1; PID:g
 R:Toman, P.D.; Ricca, G.A.; de Crombrughe, B.
 Nucleic Acids Res. 16, 7201, 1988
 A/Title: Nucleotide sequence of a cDNA coding for the amino-terminal region of human pre
 A/Reference number: S01726; MUID:88303360; PMID:3405773
 A/Accession: S01726

A/Molecule type: mRNA
 A/Residues: 1-170 <TOM>
 A/Cross-references: UNIPARC:UPI000016A706; EMBL:X07240; NID:g30060; PIDD:CAA30229.1; PID:
 A/Note: the authors translated the codon CAG for residue 154 as His
 R:Janeczko, R.A.; Ramirez, F.
 Nucleic Acids Res. 17, 6742, 1989
 A/Title: Nucleotide and amino acid sequences of the entire human alpha-1 (III) collagen.
 A/Reference number: S04887; MUID:89386015; PMID:2780304
 A/Accession: S04887
 A/Molecule type: mRNA
 A/Residues: 149-163,'G',164-240,'D',242-471,'D',473-487,'L',489,'S',491-613,'Y',615-634,'
 A/Cross-references: UNIPARC:UPI000016A61C; EMBL:X15332; NID:g29545; PIDD:CAA33387.1; PID:

A>Note: the authors' translation of residues 905-932 is inconsistent with the nucleotide
 R:Sever, J.M.; Kang, A.H.
 Biochemistry 16, 1158-1164, 1977
 A>Title: Covalent structure of collagen: amino acid sequence of cyanogen bromide peptide
 A:Reference number: A90399; PMID:77134724; PMID:557335
 A:Accession: A90399
 A:Molecule type: protein
 A:Residues: 'V', 169-225, 229-232, 'P', 234-292, 'D', 294-398 <SEY1>
 A:Cross-references: UNIPARC:UPI0000173881
 A:Experimental source: liver
 A>Note: sequence corrected by A94562; attachment of 2-O-alpha-D-glucosyl-O-beta-D-galact
 R:Sever, J.M.
 submitted to the Atlas, December 1977
 A:Reference number: A94562
 A:Accession: A94562
 A:Molecule type: protein
 A:Residues: 'V', 169-225, 229-277, 'A', 279-292, 'D', 294, 'S', 296-398 <SEY2>
 A:Cross-references: UNIPARC:UPI0000173882
 A:Experimental source: liver
 A>Note: author submitted corrections to A90399
 R:Milwicz, D.M.; Wiltz, A.M.; Smith, A.C.; Manchester, D.K.; Waldestein, G.; Byers, P.H.
 Am. J. Hum. Genet. 53, 62-70, 1993
 A>Title: Parental somatic and germ-line mosaicism for a multiexon deletion with unusual
 ispring.
 A:Reference number: 151868; PMID:93304430; PMID:8317500
 A:Accession: 151868
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 186-194 <MTL>
 A:Cross-references: UNIPARC:UPI000000B14; GB:S62925; NID:9386425; PIDN:AAJ13937.1; PID:
 R:Chido, A.A.; Sillence, D.O.; Cole, W.G.; Bateman, J.F.
 Biochem. J. 311, 939-943, 1995
 A>Title: Abnormal type III collagen produced by an exon-17-skipping mutation of the COL3
 A:Reference number: 559511; PMID:96067614; PMID:7487954
 A:Accession: 559511
 A:Molecule type: mRNA
 A:Residues: 302-423 <CHI>
 A:Cross-references: UNIPARC:UPI0000173883; GB:S79877; NID:91195576; PIDN:AA35615.1; PID:
 R:Sever, J.M.; Kang, A.H.
 Biochemistry 17, 3404-3411, 1978
 A>Title: Covalent structure of collagen: amino acid sequence of five consecutive CNBr pe
 A:Reference number: A90414; PMID:79000343; PMID:687591
 A:Accession: A90414
 A:Molecule type: protein
 A:Residues: 399-675, 'N', 677-727 <SEY3>
 A:Cross-references: UNIPARC:UPI0000173884
 A:Experimental source: liver
 R:Lee, B.; Vitale, E.; Superfi-Furga, A.; Steinmann, B.; Ramirez, F.
 J. Biol. Chem. 266, 5256-5259, 1991
 A>Title: G to T transversion at position +5 of a splice donor site causes skipping of th
 A:Reference number: 155349; PMID:9116161; PMID:1672129
 A:Accession: 155349
 A:Molecule type: protein
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 537-605 <LEE>
 A:Cross-references: UNIPARC:UPI000000042; GB:M59312; NID:9180815; PIDN:AA52041.1; PID:
 R:Sever, J.M.; Mainard, C.; Kang, A.H.
 Biochemistry 19, 1583-1589, 1980
 A>Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB5 from ty
 A:Reference number: A90438; PMID:80198282; PMID:6246925
 A:Accession: A90438
 A:Molecule type: protein
 A:Residues: 728-895, 'A', 897-964 <SEY4>
 A:Cross-references: UNIPARC:UPI0000173885
 A:Experimental source: liver
 R:Cole, W.G.; Chido, A.A.; Lemande, S.R.; Janeczko, R.; Ramirez, F.; Dahl, H.H.M.; Chan
 J. Biol. Chem. 265, 17070-17077, 1990
 A>Title: A base substitution at a splice site in the COL3A1 gene causes exon skipping an
 A:Reference number: A93303; PMID:91009133; PMID:2145268
 A:Accession: A93303
 A:Molecule type: mRNA
 A:Residues: 861-1015 <COL>
 A:Cross-references: UNIPARC:UPI00000004A1; GB:J05617; GB:M55603; GB:M59227; NID:9180878;

A>Note: a mutant sequence with 942-977 spliced out from a patient with Ehlers-Danlos syn
 R:Mankoo, B.S.; Dalgleish, R.
 Nucleic Acids Res. 16, 2337, 1988
 A>Title: Human pro alpha1(III) collagen: cDNA sequence for the 3' end.
 A:Reference number: 502119; PMID:88189827; PMID:3357782
 A:Accession: 502119
 A:Status: translation not shown
 A:Molecule type: mRNA
 A:Residues: 950-1018, 'Y', 1020-1183, 'S', 1185-1466 <MAN>
 A:Cross-references: UNIPARC:UPI0000173886; EMBL:X06700; NID:930053; PIDN:CAA29886.1; PID:
 R:Sever, J.M.; Kang, A.H.
 Biochemistry 20, 2621-2627, 1981
 A>Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB9 from ty
 A:Reference number: A90446; PMID:81208139; PMID:7016180
 A:Accession: A90446
 A:Molecule type: protein
 A:Residues: 965-979, 'A', 981-984, 'PS', 987, 'QN', 990-1096, 'P', 1098-1152, 'AT', 1155, 'S', 1157-
 A:Cross-references: UNIPARC:UPI0000173887
 A:Experimental source: liver
 R:Loidi, H.R.; Brinker, J.M.; May, M.; Phlajant, T.; Morrow, S.; Rosenbloom, J.; Mye
 Nucleic Acids Res. 12, 9383-9394, 1984
 A>Title: Molecular cloning and carboxyl-propeptide analysis of human type III procollagen
 A:Reference number: A93551; PMID:85087944; PMID:6096827
 A:Accession: A93551
 A:Molecule type: mRNA
 A:Residues: 1065-1155, 'P', 1157-1466 <LOT>
 A:Cross-references: UNIPARC:UPI0000173888; EMBL:X01655; EMBL:X01742; NID:929584; PIDN:CA
 R:Marklin, M.; Dalgleish, R.; Kluge-Beckmann, B.; Renard, S.I.; Tolstoev, P.; Brant
 Biochemistry 25, 1408-1413, 1986
 A>Title: Human type III collagen gene expression is coordinately modulated with the type
 A:Reference number: 152393; PMID:86187804; PMID:3754462
 A:Accession: 152393
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1161-1200 <MIS>
 A:Cross-references: UNIPARC:UPI000016A6B5; GB:M13146; NID:9180415; PIDN:AA52003.1; PID:
 R:Manuel, B.S.; Camizaro, L.A.; Sever, J.M.; Myers, J.C.
 Proc. Natl. Acad. Sci. U.S.A. 82, 3385-3389, 1985
 A>Title: Isolation of cDNA and genomic clones encoding human pro-alpha1(III) collagen. Pe
 A:Reference number: A92516; PMID:85157600; PMID:2579949
 A:Accession: A92516
 A:Molecule type: DNA
 A:Residues: 1176-1240, 'V', 1242-1356, 'P', 1358-1466 <CHD>
 A:Cross-references: UNIPARC:UPI0000173889; GB:M10615; GB:M10793; GB:M10794; GB:M10795; G
 A:Experimental source: liver
 A>Note: the authors translated the codon TTC for residue 1057 as Tyr; the codons given fo
 action
 C:comment: Prolines and lysines at the third position of the tripeptide repeating unit (C
 3-hydroxylated. About 15% of the lysines are 5-hydroxylated and some are subsequently O-
 C:genetics:
 A:Gene: GDB:COL3A1
 A:Cross-references: GDB:118729; OMIM:120180
 A:Map position: 2q31-2q31
 A:Introns: 27/1, 94/3, 111/3, 149/3, 176/3, 554/3, 587/3, 1175/3, 1275/1, 1337/3, 1418/3
 A>Note: the list of introns is incomplete; defects in this gene can result in Ehlers-Dan
 C:complex: type III collagen is a homotrimer of monomers initially linked by disulfide bo
 er of their length, is formed with desmosine cross-links made from lysine and allylsine re
 C:Function:
 A:Description: structural component of extracellular fibrous polymer that maintains intec
 C:superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
 C:keywords: coiled coil; Ehlers-Danlos syndrome; extracellular matrix; glycoprotein; hyd
 F/1-23/Dominant: signal sequence #starts predicted <SID>
 F/24-153/Dominant: amino-terminal propeptide #repeats predicted <PRO>
 F/31-91/Dominant: von Willebrand factor type C repeat homology <VWC>

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 1, 2005, 08:15:46 ; Search time 228 Seconds
(without alignments)
504.391 Million cell updates/sec

Title: US-09-924-340-58
Perfect score: 952
Sequence: 1 MGPFGFKGTGHGFLPGPKG.....GAMPMEQGYPMKTKMGPPG 163

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues
Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : UniProt_05.80:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	840.5	88.3	1463	2	059F89_HUMAN
2	840.5	88.3	1603	1	COGAL_HUMAN
3	800.5	84.1	575	2	08CIF9_MOUSE
4	794.5	83.5	415	2	0568Y4_RAT
5	777	81.6	1560	2	08BLX7_MOUSE
6	620	65.1	218	2	071RG9_HUMAN
7	558	58.6	150	2	09CZS2_MOUSE
8	407	42.8	925	2	04SIU4_TETNG
9	391	41.1	799	2	08BNS7_MOUSE
10	391	41.1	1691	2	09ESQ2_MOUSE
11	391	41.1	1691	2	0632W6_MOUSE
12	390	41.0	1685	1	CO4A5_HUMAN
13	390	41.0	1685	1	09NIB7_HUMAN
14	387	40.7	1146	1	COHAL_CHICK
15	387	40.7	1453	1	COHAL_CHICK
16	385	40.4	1497	1	COHAL_HUMAN
17	384	40.3	1549	2	060444_CRIGR
18	384	40.3	1748	2	0591P1_PIG
19	383.5	40.3	1840	2	CO5A1_HUMAN
20	383.5	40.3	1840	2	0591P3_PIG
21	383	40.2	355	2	05YFP3_GVIRU
22	383	40.2	674	1	COA11_BOVIN
23	382.5	40.2	591	2	096HC0_HUMAN
24	382.5	40.2	848	2	04S270_TETNG
25	382.5	40.2	1049	1	CO3A1_BOVIN
26	382.5	40.2	1415	2	05JVJ6_HUMAN
27	382.5	40.2	1792	2	059SE7_HUMAN
28	382.5	40.2	1838	2	015094_HUMAN
29	382.5	40.2	1838	2	05SUX4_HUMAN
30	382.5	40.2	1838	2	088207_MOUSE
31	382.5	40.2	1840	2	060467_CRILLO

32	382.5	40.2	1840	2	09J103_RAT	09J103_rattus norv
33	382	40.1	1694	2	08HYC1_CANFA	08hy1 canis famil
34	382	40.1	1691	1	CO4A5_CANFA	028247 canis famil
35	381.5	40.1	657	2	04VAP9_HUMAN	04vap9 homo sapien
36	381.5	40.1	703	1	CO8A2_HUMAN	05067 homo sapien
37	381.5	40.1	703	2	05JVJ1_HUMAN	05jv1 homo sapien
38	381	40.0	367	2	05GAF3_GVIRU	05gaf3 grouper iri
39	380.5	40.0	1470	1	COHAL_MOUSE	007553 mus musculu
40	380.5	40.0	1835	2	09IAU4_CHICK	09ia4 gallus gall
41	380	39.9	1806	2	05VM31_HUMAN	05vt31 homo sapien
42	379	39.8	675	2	09N178_PIG	09n178 sus scrofa
43	378.5	39.8	1269	2	07T227_CHICK	07t227 gallus gall
44	378.5	39.8	1658	2	059GD4_HUMAN	059gd4 homo sapien
45	378.5	39.8	1745	1	CO5A3_HUMAN	P25940 homo sapien
46	378	39.7	1445	2	093251_RANCA	093251 rana catesb
47	377.5	39.7	888	2	090796_CHICK	090796 gallus gall
48	377	39.6	568	2	08CD80_MOUSE	08cd80 mus musculu
49	377	39.6	1163	2	08N6U4_HUMAN	08n6u4 homo sapien
50	377	39.6	1463	2	05PQT6_RAT	05pqt6 rattus norv
51	377	39.6	1463	2	CO3A1_HUMAN	P02461 homo sapien
52	377	39.6	1466	2	053S91_HUMAN	053s91 homo sapien
53	377	39.6	1466	2	0541P8_HUMAN	0541p8 homo sapien
54	376.5	39.5	1449	2	06NZ15_BRARE	06nz15 brachydanio
55	376.5	39.5	1487	2	077753_CANFA	077753 canis famil
56	376.5	39.5	1690	1	CO4A4_HUMAN	P53420 homo sapien
57	376.5	39.5	1690	1	053WRI_HUMAN	053wr1 homo sapien
58	376	39.5	1652	1	CO3A1_CHICK	P12105 gallus gall
59	375	39.5	1450	2	09YIB4_CYNPY	09yib4 cynops pyrr
60	375.5	39.4	1449	2	06PE19_BRARE	06pe19 brachydanio
61	374.5	39.3	585	2	08OV57_MOUSE	08ov57 mus musculu
62	374	39.3	680	2	09DOD2_MOUSE	09dod2 m mus muscu
63	374	39.3	1464	1	CO3A1_MOUSE	P08121 mus musculu
64	374	39.3	1464	2	08BKX2_MOUSE	08bkx2 mus musculu
65	374	39.3	1464	2	08BLW4_MOUSE	08blw4 mus musculu
66	374	39.3	1464	2	07TT32_MOUSE	07tt32 mus musculu
67	373.5	39.2	1467	2	05DTG2_MOUSE	05dtg2 mus musculu
68	373.5	39.2	1804	1	COBA1_MOUSE	061245 mus musculu
69	373.5	39.2	1804	1	08OWR4_MOUSE	08owr4 mus musculu
70	373	39.2	890	2	077087_PANNE	077087 alvinella p
71	373	39.2	1770	2	07Q1V4_ANOCA	07q1v4 anopheles g
72	372.5	39.1	280	2	091014_CHICK	091014 gallus gall
73	372.5	39.1	680	1	COA11_HUMAN	003622 homo sapien
74	372.5	39.1	739	2	070575_MOUSE	070575 mus musculu
75	372.5	39.1	977	2	04SEP8_TETNG	04sep8 tetraodon n
76	372.5	39.1	1420	2	090W37_CHICK	090w37 gallus gall
77	372	39.1	662	2	08BYX3_MOUSE	08byx3 mus musculu
78	372	39.1	680	2	08BSX1_MOUSE	08bsx1 mus musculu
79	372	39.1	920	1	CO9A1_CHICK	P12106 gallus gall
80	372	39.1	1069	2	06LAN8_HUMAN	06lan8 homo sapien
81	372	39.1	1449	2	0640B2_XENTR	0640b2 xenopus tro
82	372	39.1	1461	2	076045_HUMAN	076045 homo sapien
83	372	39.1	1464	1	COA11_HUMAN	P02442 homo sapien
84	372	39.1	1464	2	08N473_HUMAN	08n473 homo sapien
85	372	39.1	1467	2	059F64_HUMAN	059f64 homo sapien
86	371.5	39.0	730	2	026052_PARLI	026052 paracentroc
87	371.5	39.0	1017	2	059HB5_HUMAN	059hb5 homo sapien
88	371.5	39.0	1806	1	COBA1_HUMAN	P12107 homo sapien
89	371	39.0	636	1	CO3A1_RAT	P13919 rattus norv
90	371	39.0	779	1	COA11_BOVIN	P02433 bos taurus
91	371	39.0	1460	1	COA11_CANFA	09x970 canis famil
92	371	39.0	1631	2	059700_CANFA	059700 canis famil
93	371	39.0	1707	1	CO4A2_MOUSE	P08122 mus musculu
94	370.5	38.9	826	2	08KON2_MOUSE	08kon2 mus musculu
95	370.5	38.9	886	2	08CEP7_MOUSE	08cep7 mus musculu
96	370.5	38.9	886	2	08OYV3_MOUSE	08oyv3 mus musculu
97	370.5	38.9	1419	2	08OX38_MOUSE	08ox38 mus musculu
98	370.5	38.9	1419	2	063123_RAT	063123 rattus norv
99	370.5	38.9	1442	2	062031_MOUSE	062031 mus musculu
100	370.5	38.9	1442	2	062033_MOUSE	062033 mus musculu

ALIGNMENTS

RESULT 1
 OS9F89 HUMAN PRELIMINARY; PRT; 1463 AA.
 AC OS9F89; 10-MAY-2005 (TRENBLREL. 30, Created)
 DT 10-MAY-2005 (TRENBLREL. 30, Last sequence update)
 DT 10-MAY-2005 (TRENBLREL. 30, Last annotation update)
 DE Alpha 1 type XVI collagen variant (Fragment).
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae;
 OC Homo.
 OC NCBI_Taxid=9606;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Skin;
 RA Totoki Y., Toyoda A., Takeda T., Sakaki Y., Tanaka A., Yokoyama S.,
 RA Ohara O., Nagase T., Kikuno F.R.;
 RT "None Title";
 RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB209571; BMD92808.1; -; mRNA.
 DR GO; GO:0005737; C:cytoplasm; IEA.
 DR GO; GO:0005578; C:extracellular matrix (sensu Metazoa); IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR GO; GO:0007155; P:cell adhesion; IEA.
 DR GO; GO:006817; P:phosphate transport; IEA.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR003129; Laminin_G_TSP_N.
 DR Pfam; PF01391; Collagen; 14.
 DR SMART; SM00210; TSPN; 1.
 KW Collagen; Extracellular matrix.
 FT NON TER 1
 SQ SEQUENCE 1463 AA; 142798 MW; 486F0B10CA93785C CRC64;

Query Match 88.3%; Score 840.5; DB 2; Length 1463;
 Best Local Similarity 44.5%; Pred. No. 3.6e-42;
 Matches 163; Conservative 0; Mismatches 0; Indels 203; Gaps 1;

1 MBPPGKGTGHPGLPGPKGDKPPGSTGRPGAEGPGAMGPGPGPGHVPGP 60
 1098 MBPPGKGTGHPGLPGPKGDKPPGSTGRPGAEGPGAMGPGPGPGHVPGP 1157
 61 PQQPGAGISAVGLKGRGATGERGLAGPGQ----- 92
 1158 PQQPGAGISAVGLKGRGATGERGLAGPGQPGPPGHPGPEPETDGAAGKSPGKQ 1217
 93 ----- 92
 1218 GFYPPGPKGDPCGAAGKQAGEKGRAGMPGPGKSGMGPVGPQPGAGERHPGAAPS 1277
 93 ----- 92
 1278 GSPGLPGVSGMGDMVYDEIKRFIRQEI11KMFDEMAVYTSRQPMEMAAAQGRPGPP 1337
 93 -----PGPG 97
 1338 GKDGAAGRGARQSPGLPQIGREGROGLPGVRLPOTGKEKDIGIAGENGGLPGPG 1397
 98 PQQPGYGMKATGPMGQGGIPGIPGPPGPMGPGKAGHCNPSDCGAMPMEQYPPMK 157
 1398 PQQPGYGMKATGPMGQGGIPGIPGPPGPMGPGKAGHCNPSDCGAMPMEQYPPMK 1457
 158 MKGPPG 163
 1458 MKGPPG 1463

RESULT 2
 COGAL HUMAN STANDARD; PRT; 1603 AA.
 AC COGAL; 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 10-MAY-2005 (Rel. 47, Last annotation update)
 DE Collagen alpha 1(XVI) chain precursor.
 GN Name=COL16A1;
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae;
 OC Homo.
 OC NCBI_Taxid=9606;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA MEDLINE=9235339; PubMed=1631157;
 RA Pan T.C., Zhang R.Z., Mattei M.-G., Timpl R., Chu M.-L.,
 RT "Cloning and chromosomal location of human alpha 1(XVI) collagen.";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:6565-6569(1992).
 RN [2]
 RP NUCLEOTIDE SEQUENCE OF 418-1603.
 RC TISSUE=Placenta;
 RX MEDLINE=93203161; PubMed=1284248;
 RA Yamaguchi N., Kimura S., McBride O.W., Hori H., Yamada Y.,
 RA Kanamori T., Yamakoshi H., Nagai Y.;
 RT "Molecular cloning and partial characterization of a novel collagen
 chain, alpha 1(XVI), consisting of repetitive collagenous domains and
 cysteine-containing non-collagenous segments.";
 RL J. Biochem. 112:856-863(1992).
 CC -1- FUNCTION: The numerous interruptions in the triple helix may make
 this molecule either elastic or flexible.
 CC -1- TISSUE SPECIFICITY: In the placenta, where it is found in the
 amnion, it is found in an acellular, relatively dense layer of a
 complex network of reticular fibers. Also located to a fibroblast
 layer beneath this dense layer. Exists in tissues in association
 with other types of collagen.
 CC -1- DEVELOPMENTAL STAGE: Transiently elevated expression during
 gestation, and decrease at term.
 CC -1- DOMAIN: This sequence defines eighteen different domains, nine
 triple-helical domains (COL9 to COL1) and ten nontriple-helical
 domains (NC10 to NC1).
 CC -1- PTM: Prolines at the third position of the tripeptide repeating
 unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC -1- SIMILARITY: Belongs to the fibril-associated collagens with
 interrupted helices (FACIT) family.
 CC -1- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use as long as its content is in no way modified and this statement is not
 removed.
 CC EMBL; M92642; AAA58427.1; -; mRNA.
 CC EMBL; S57132; AAB25797.1; -; mRNA.
 DR PIR; S23810; S23810.
 DR Ensembl; ENSG0000084636; Homo sapiens.
 DR HGNC; HGNC:2193; COL16A1.
 DR MIM; 120326; -.
 DR GO; GO:0005597; C:collagen type XVI; TAS.
 DR GO; GO:0007565; P:pregnancy; TAS.
 DR InterPro; IPR008161; C1g_helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR003129; Laminin_G_TSP_N.
 DR Pfam; PF01391; Collagen; 16.
 DR Prodom; PD0000007; C1g_helix; 1.
 DR SMART; SM00210; TSPN; 1.
 KW Collagen; Extracellular matrix; Hydroxylation; Repeat; Signal;
 KW Structural protein.
 FT SIGNAL 1 21 Potential.
 FT CHAIN 22 1603 Collagen alpha 1(XVI) chain.
 FT DOMAIN 50 231 TSP N-terminal.
 FT REGION 222 374 Nonhelical region 10 (NC10).
 FT REGION 375 505 Triple-helical region 9 (COL9) with 3
 FT imperfections.
 FT REGION 506 520 Nonhelical region 9 (NC9).

FT REGION 521 554 Triple-helical region 8 (COL8) with 1
 FT REGION 555 571 Imperfection.
 FT REGION 572 630 Nonhelical region 8 (NC8).
 FT REGION 631 651 Triple-helical region 7 (COL7) with 1
 FT REGION 652 722 Imperfection.
 FT REGION 723 737 Nonhelical region 7 (NC7).
 FT REGION 738 875 Triple-helical region 6 (COL6) with 1
 FT REGION 876 886 Imperfection.
 FT REGION 887 938 Nonhelical region 5 (NC5).
 FT REGION 939 972 Triple-helical region 4 (COL4) with 2
 FT REGION 973 987 Imperfection.
 FT REGION 988 1010 Nonhelical region 3 (NC3).
 FT REGION 1011 1432 Triple-helical region 2 (COL2) with 2
 FT REGION 1433 1471 Imperfection.
 FT REGION 1472 1577 Nonhelical region 1 (NC1).
 FT REGION 1578 1603 Triple-helical region 1 (NC1) with 2
 FT CONFLICT 418 420 RDA -> GGR (in Ref. 2).
 FT CONFLICT 537 537 R -> P (in Ref. 2).
 FT CONFLICT 1160 1160 T -> P (in Ref. 2).
 FT CONFLICT 1163 1163 T -> P (in Ref. 2).
 FT CONFLICT 1165 1165 S -> P (in Ref. 2).
 SQ SEQUENCE 1603 AA; 157693 MW; E27D9A1D4E598A37 CRC64;

Query Match 88.3%; Score 840.5; DB 1; Length 1603;
 Best Local Similarity 44.5%; Pred. No. 3,9e-42;
 Matches 163; Conservative 0; Mismatches 0; Indels 203; Gaps 1;

QY 1 MGPRGKGTGHPGLPGPKDCKPSPGSTRPAGBEGPAGMPQGRPGPHVGP 60
 DB 1238 MGPRGKGTGHPGLPGPKDCKPSPGSTRPAGBEGPAGMPQGRPGPHVGP 1297
 QY 61 PGQPGAGTAVGLKDRGATGSRGLAGPQ----- 92
 DB 1298 PGQPGAGTAVGLKDRGATGSRGLAGPQPGPHGPPEPPTGDAAGKEGPPGKQ 1357
 QY 93 ----- 92
 DB 1358 GFYGPGRPKDPAAGQKQAGBKGRAGMPPGKSGSMGPVPPGABGRHGPAGPS 1417
 QY 93 ----- 92
 DB 1418 GSPGLPGVDSMGDMVNDIKRIFRQELIKMFERMAYVYTSHPFMEVAAAPGRPGP 1477
 QY 93 -----PGPPG 97
 DB 1478 GKDGAAPRGAPGSPGLPGQIGREGROGLPGVGLPOTKGEKDDIGIAGENGGLPGPPG 1537
 QY 98 PGQPGYGMGATGPMGQGGIPGIPGPPGMPQPGKAGHCNPSSDCGAMPMEQOYPPMKT 157
 DB 1538 PGQPGYGMGATGPMGQGGIPGIPGPPGMPQPGKAGHCNPSSDCGAMPMEQOYPPMKT 1597
 QY 158 MKGPPG 163
 DB 1598 MKGPPG 1603

RESULT 3
 Q8CIF9_MOUSE PRELIMINARY; PRT; 575 AA.
 AC Q8CIF9;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Coll16a1 protein (Fragment).
 GN Name=Coll16a1;
 OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
 OC Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP NCLEOTIDE SEQUENCE.
 RC STRAIN=FVB/N; TISSUE=Mammary tumor. C3;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Peingold E.A., Grouse L.H., Derge J.G.,
 RA Klauener R.D., Collins F.S., Wagner L., Stammen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusik A., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W.,
 RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP NCLEOTIDE SEQUENCE.
 RC STRAIN=FVB/N; TISSUE=Mammary tumor. C3;
 RA Strausberg R.;
 RL Submitted (FEF-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC023940; AAH23940.1; -; mRNA.
 DR Ensembl: ENSMUSG00000040690; Mus musculus.
 DR MGI: MGI:1095396; Coll16a1.
 DR GO: GO:0005737; Cytoplasm; IEA.
 DR GO: GO:0006817; PIP2 transport; IEA.
 DR InterPro: IPR008161; C1g helix.
 DR InterPro: IPR008160; Collagen.
 DR Pfam: PF01391; Collagen; 9.
 DR ProDom: PD000007; C1g_helix; 1.
 FT NON TER 1 1
 SQ SEQUENCE 575 AA; 54847 MW; 0B542D0D4086A313 CRC64;

Query Match 84.1%; Score 800.5; DB 2; Length 575;
 Best Local Similarity 42.1%; Pred. No. 4e-40;
 Matches 154; Conservative 4; Mismatches 5; Indels 203; Gaps 1;

QY 1 MGPRGKGTGHPGLPGPKDCKPSPGSTRPAGBEGPAGMPQGRPGPHVGP 60
 DB 210 MGPRGKGTGHPGLPGPKDCKPSPGSTRPAGBEGPAGMPQGRPGPHVGP 269
 QY 61 PGQPGAGTAVGLKDRGATGSRGLAGPQ----- 92
 DB 270 PGQPGAGTAVGLKDRGATGSRGLAGPQPGPHGPPEPPTGDAAGKEGPPGKQ 329
 QY 93 ----- 92
 DB 330 GLYGPGRPKDPAAGQKQAGBKGRAGMPPGKSGSMGPVPPGABGRHGPAGPS 389
 QY 93 ----- 92
 DB 390 GNPGLPGVDSMGDMVNDIKRIFRQELIKMFERMAYVYTSHPFMEVAAAPGRPGP 449
 QY 93 -----PGPPG 97
 DB 450 GKDGAAPRGAPGSPGLPGQIGREGROGLPGVGLPOTKGEKDDIGIAGENGGLPGPPG 509
 QY 98 PGQPGYGMGATGPMGQGGIPGIPGPPGMPQPGKAGHCNPSSDCGAMPMEQOYPPMKT 157
 DB 510 PGQPGYGMGATGPMGQGGIPGIPGPPGMPQPGKAGHCNPSSDCGAMPMEQOYPPMKT 569

```

OY      158 MKGPG 163
      |||||
Db      570 MKGPG 575

RESULT 4
O568Y4 RAT PRELIMINARY; PRT; 415 AA.
ID O568Y4
AC O568Y4
DT 10-MAY-2005 (TRENBLrel. 30, Last sequence update)
DT 10-MAY-2005 (TRENBLrel. 30, Last annotation update)
DE Col16a1 predicted protein.
DE Col16a1 predicted protein.
DE Name=Col16a1 predicted;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Rattus.
OC NCBI_TaxId=10116;
[1]
RN NUCLEOTIDE SEQUENCE.
RC MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Scheffer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant J.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mallaby S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Hellon E., Kettman W., Madan A., Rodriguez Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalek U., Smallos D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
[2]
RN NUCLEOTIDE SEQUENCE.
RC TISSUE=Brain;
RG NIH MGC Project;
RL Submitted (APR-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC092654; AAH92654.1; -; mRNA.
DR GO; GO:0005737; Cytoplasm; IEA.
DR GO; GO:0006817; Phosphate transport; IEA.
KW Collagen.
SQ SEQUENCE 415 AA; 39879 MW; 693F7D948237D1AA CRC64;

Query Match 83.5%; Score 794.5; DB 2; Length 415;
Best Local Similarity 41.5%; Pred. No. 7e-40;
Matches 152; Conservative 7; Mismatches 4; Indels 203; Gaps 1;

OY      1 MKGPGKXTGHPGLPQPKDCKPFGPSTGTPGAGGANGPQGRPGPPHVPDP 60
      |||||
Db      50 MGPPGKXTGHPGLPQPKDCKPFGPSTGTPGAGGANGPQGRPGPPHVPDP 109
      |||||
OY      61 PGOPGAGISAVGLKSDRGATGGRGLAGLPGQ----- 92
      |||||
Db      110 PGOPGPGGLSTWGLKSDRGSPGRGLAGLPGQPTGHPGPBPSPGSDGAGKEGPPGKQ 169
      |||||
OY      93 ----- 92
      |||||
Db      170 GLYGPGRKDPGAPGAGKGRSGMGPGRSGSGMGPICPPGAPGRGHPGSPGPA 229
      |||||
OY      93 ----- 92
      |||||

```

```

Db      230 GNPGLPGLPGSMGVNVTDDIKRFIRQELIKLFDERMAYYSRMQPMVEVAAAPGRPGP 289
      |||||
OY      93 ----- PGPG 97
      |||||
Db      290 GKDGA-PGRPGAPGSPGLPQIGREGRGQGLPGWRGLPGTKGDKGIDGVGIAGENGLPGPG 349
      |||||
OY      98 PGPPGPGYGMGATGPMGQGGIPGIPGPPGPMQPKAGHCNPSDDCGAMPMEQOYPPMT 157
      |||||
Db      350 PGPPGPGYGMGATGPMGQGGIPGIPGPPGPMQPKAGHCNPSDDCGAMPMEQOYPPMT 409
      |||||
OY      158 MKGPG 163
      |||||
Db      410 MKGPG 415

RESULT 5
O8BLX7 MOUSE
ID O8BLX7_MOUSE PRELIMINARY; PRT; 1580 AA.
AC O8BLX7;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE Mus musculus adult male aorta and vein cDNA, RIKEN full-length
DE enriched library, clone:AS3052M23 product:COLLAGEN ALPHA 1(XVI) CHAIN
DE homolog.
GN Name=Col16a1;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
OC NCBI_TaxId=10090;
[1]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Aorta and vein;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.,
RT "High-efficiency full-length cDNA cloning."
RT Meth. Enzymol. 303:19-44 (1999).
[2]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Aorta and vein;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Aikawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batilov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Koehliwa H.,
RA Kueni P., Lewis S., Matsuo T., Nikaido I., Pesole G., Quackenbush J.,
RA Schirini L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli U., Mombarts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitteker C., Wilming L.,
RA Wyszynski A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.,
RT "Functional annotation of a full-length mouse cDNA collection."
RT Nature 409:685-690 (2001).
[3]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Aorta and vein;
RA The FANTOM Consortium.
RA "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RT Nature 420:563-573 (2002).
[4]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Aorta and vein;

```


DR ProDom; PD000007; C1g_helix; 2.
 DR SMART; PD003923; Procollagen4_C; 2.
 DR SMART; SM00111; C4; 2.
 KW Collagen.
 FT NON_TER
 SQ SEQUENCE 1 1 799 AA; 77889 MW; C517CFACF15706DC CRC64;

Query Match 41.1%; Score 391; DB 2; Length 799;
 Best Local Similarity 49.7%; Pred. No. 1e-15;
 Matches 82; Conservative 13; Mismatches 52; Indels 18; Gaps 5;

QY 2 GPPGFGKTKGHHLPKPKDCKPGRPGSTGRBAGEGCGANGPQGRPGPHVGPSPGP 61
 DB 105 GPPGSLGGQSPGLPCKGNPGLPGPPGLTGPGRKGN---IGDWGFPGPQGVDPGPP 161
 QY 62 GPPGAGISAVGLKGRGATGERGLA--GLPGQPPGPQGPQGPY-GKMGATGPMQOQGI 118
 DB 162 GPPGQPG--SPGLPGQKSKGEPVSGITLPGLPGRKSGPLPGYGNPGIGSGVETGL 219
 QY 119 PGIPGPPGPMGQPGKAGHCNPSPDCFGAMPMEQOYPPMKTKMPFG 163
 DB 220 PGLPGNPGAKGQGLPG-----PFGTGLPGPKGMNGPFG 254

RESULT 10
 Q9ESQ2 MOUSE PRELIMINARY; PRT; 1691 AA.

ID Q9ESQ2 MOUSE PRELIMINARY; PRT; 1691 AA.
 AC Q9ESQ2;
 DT 01-MAR-2001 (TREMblrel. 16, Created)
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
 DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)
 DE Type IV collagen alpha 5 chain.
 GN Name=C0145;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
 OC Muridae; Murinae; Mus.
 OC NCBI_TaxId=10090;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RX MEDLINE=2036494; PubMed=10965041;
 RA Saiko K., Naito I., Seki T., Ohashi T., Kimura E., Momota R.,
 RA Kishino Y., Sado Y., Yoshioke H., Ninomiya Y.,
 RA "Differential Expression of Mouse a5(IV) and a6(IV) Collagen Genes in
 RA Epithelial Basement Membranes."
 RT J. Biochem. 128:427-434(2000).
 RL EMBL; AB041350; BAB13673.1; -; mRNA.
 DR HSSP; P02462; 1L11.
 DR SMR; Q9ESQ2; 1464-1691.
 DR Ensemble; ENSMUSG00000031274; Mus musculus.
 DR MGI; MGI:88456; C0145.
 DR GO; GO:0005604; C:basement membrane; IDA.
 DR GO; GO:0005615; C:extracellular space; TAS.
 DR InterPro; IPR008161; C1g_helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR001442; Procollagen4_C.
 DR Pfam; PF01413; C4; 2.
 DR Pfam; PF01391; Collagen; 24.
 DR ProDom; PD000007; C1g_helix; 3.
 DR ProDom; PD003923; Procollagen4_C; 2.
 DR SMART; SM00111; C4; 2.
 KW Collagen.
 SQ SEQUENCE 1691 AA; 161823 MW; 81340DF1792208FA CRC64;

Query Match 41.1%; Score 391; DB 2; Length 1691;

Best Local Similarity 49.7%; Pred. No. 1.9e-15;
 Matches 82; Conservative 13; Mismatches 52; Indels 18; Gaps 5;

QY 2 GPPGFGKTKGHHLPKPKDCKPGRPGSTGRBAGEGCGANGPQGRPGPHVGPSPGP 61
 DB 991 GPPGSLGGQSPGLPCKGNPGLPGPPGLTGPGRKGN---IGDWGFPGPQGVDPGPP 1047
 QY 62 GPPGAGISAVGLKGRGATGERGLA--GLPGQPPGPQGPQGPY-GKMGATGPMQOQGI 118

DB 1048 GPPGQPG--SPGLPGQKSKGEPVSGITLPGLPCKGNPGLPGYGNPGIGSGVETGL 1105
 QY 119 PGIPGPPGPMGQPGKAGHCNPSPDCFGAMPMEQOYPPMKTKMPFG 163
 DB 1106 PGLPGNPGAKGQGLPG-----PFGTGLPGPKGMNGPFG 1140

RESULT 11
 O632W6 MOUSE PRELIMINARY; PRT; 1691 AA.

ID O632W6;
 AC O632W6;
 DT 25-OCT-2004 (TREMblrel. 28, Created)
 DT 25-OCT-2004 (TREMblrel. 28, Last sequence update)
 DT 25-OCT-2004 (TREMblrel. 28, Last annotation update)
 DE Procollagen, type IV, alpha 5.
 GN Name=C0145;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
 OC Muroidae; Muridae; Murinae; Mus.
 OC NCBI_TaxId=10090;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=FVB/N; TISSUE=Mammary tumor. C3;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Scheffer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Cavaletto T.L., Scheetz T.E.,
 RA Brownstein M.J., Uedlin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Huijck S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahney J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Buterfield Y.S.N., Krzywinski M.J., Skalska U., Smallus D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=FVB/N; TISSUE=Mammary tumor. C3;
 RG NIH MGC Project;
 RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC082788; AAH82788.1; -; mRNA.
 DR SMR; O632W6; 1464-1691.
 DR Ensemble; ENSMUSG00000031274; Mus musculus.
 DR MGI; MGI:88456; C0145.
 DR GO; GO:0005581; C:collagen; IEA.
 DR GO; GO:0005737; C:cytoplasm; IEA.
 DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
 DR GO; GO:0006817; P:phosphate transport; IEA.
 DR InterPro; IPR008161; C1g_helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR009765; Pericardial_rpt.
 DR InterPro; IPR001442; Procollagen4_C.
 DR Pfam; PF01413; C4; 2.
 DR Pfam; PF01391; Collagen; 21.
 DR ProDom; PD007054; Pericardial_rpt; 2.
 DR ProDom; PD003923; Procollagen4_C; 2.
 DR SMART; SM00111; C4; 2.
 KW Collagen.
 SQ SEQUENCE 1691 AA; 161807 MW; E5CD42658128DD07 CRC64;

Query Match 41.1%; Score 391; DB 2; Length 1691;

Best Local Similarity 49.7%; Pred. No. 1.9e-15;
Matches 82; Conservative 13; Mismatches 52; Indels 18; Gaps 5;
QY 2 GPPFGKTKHPLPGPKGDCGKRGPRGSTRGABSEBPAMRGQRGPRGHVGPGR 61
DB 991 GQGLSGQPSPLPGPKGNPLPGPGLTGPCKN---IDMGPPGQGVDPGPP 1047
QY 62 GQGPAGISAVGKGRGATGERGLA--GLPGQPGPPGQPPGY--GRMGATGPMGQGI 118
DB 1048 GFGPGG--SPGLPGKSGKSGEPVSGISGLPLGPRGEGGLGYPGNPGLKSGVETGL 1105
QY 119 PGIPGPPGMPGQPKAGHCNPSCFGAMPMEQGYPPMKTKGPPG 163
DB 1106 PGIPGPPGAKGQGLPG-----FPGTGLPGPKGNMGPPG 1140
RESULT 12
CO4A5_HUMAN STANDARD; PRT; 1685 AA.
ID CO4A5_HUMAN P29400; Q16006; Q16126; O6LD84;
AC P29400; Q16006; Q16126; O6LD84;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Collagen alpha 5 (IV) chain precursor.
GN Name=COL4A5;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA MEDLINE=94165049; PubMed=8120014;
RT Zhou J., Leinonen A., Tryggvason K.;
RT "Structure of the human type IV collagen COL4A5 gene";
RL J. Biol. Chem. 269:6608-6614(1994).
RN [2]
RP NUCLEOTIDE SEQUENCE OF 1-910, AND VARIANT AS CYS-521.
RC TISSUE=Kidney;
RA MEDLINE=92316923; PubMed=1352287;
RT Zhou J., Hertz J.M., Leinonen A., Tryggvason K.;
RT "Complete amino acid sequence of the human alpha 5 (IV) collagen chain
and identification of a single-base mutation in exon 23 converting
glycine 521 in the collagenous domain to cysteine in an Alport
syndrome patient";
RL J. Biol. Chem. 267:12475-12481(1992).
RN [3]
RP NUCLEOTIDE SEQUENCE OF 85-1685.
RC TISSUE=Placenta;
RA MEDLINE=90337990; PubMed=2380186;
RT Pihlajaniemi T., Pohjolainen E.R., Myers J.C.;
RT "Complete primary structure of the triple-helical region and the
carboxyl-terminal domain of a new type IV collagen chain, alpha
5(IV).";
RL J. Biol. Chem. 265:13758-13766(1990).
RN [4]
RP NUCLEOTIDE SEQUENCE OF 924-1685.
RA MEDLINE=91169491; PubMed=2004755;
RT Zhou J., Hostikka S.L., Chow L.T., Tryggvason K.;
RT "Characterization of the 3' half of the human type IV collagen alpha 5
gene that is affected in the Alport syndrome";
RL Genomics 9:1-9(1991).
RN [5]
RP NUCLEOTIDE SEQUENCE OF 914-1685.
RA MEDLINE=90160375; PubMed=1689491;
RT Hostikka S.L., Eddy R.L., Byers M.G., Hoeyhtyae M., Shows T.B.,
RT Tryggvason K.;
RT "Identification of a distinct type IV collagen alpha chain with
restricted kidney distribution and assignment of its gene to the locus
of X chromosome-linked Alport syndrome";
RL Proc. Natl. Acad. Sci. U.S.A. 87:1606-1610(1990).
RN [6]
RP NUCLEOTIDE SEQUENCE OF 1442-1471.

RX MEDLINE=90252791; PubMed=2339699;
RA Myers J.C., Jones T.A., Pohjolainen E.R., Kadri A.S., Goddard A.D.,
RA Sheer D., Solomon E., Pihlajaniemi T.;
RT "Molecular cloning of alpha 5(IV) collagen and assignment of the gene
to the region of the X chromosome containing the Alport syndrome
locus";
RL Am. J. Hum. Genet. 46:1024-1033(1990).
RN [7]
RP NUCLEOTIDE SEQUENCE OF 1-20.
RA Guo C., van Damme B., Vanrenterghem Y., Devriendt K., Cassiman J.-J.,
RA Marynen P.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
RN [8]
RP NUCLEOTIDE SEQUENCE OF 1258-1270 (ISOFORM 2).
RX MEDLINE=94133540; PubMed=8301933;
RA Guo C., van Damme B., van Damme-Lombaerts R., van den Berghe H.,
RA Cassiman J.-J., Marynen P.;
RT "Differential splicing of COL4A5 mRNA in kidney and white blood cells:
a complex mutation in the COL4A5 gene of an Alport patient deletes the
NC1 domain";
RL Kidney Int. 44:1316-1321(1993).
RN [9]
RP NUCLEOTIDE SEQUENCE OF 1589-1598 AND 1677-1685, AND VARIANTS AS
RP 1597-TYR--THR-1685 DEL AND 1679-GLU--THR-1685 DEL.
RX PubMed=7853788;
RA Nakazato H., Hattori S., Ushijima T., Matsura T., Koitabashi Y.,
RA Takada T., Yoshioke K., Endo F., Matsuda I.;
RT "Mutations in the COL4A5 gene in Alport syndrome: a possible mutation
in primordial germ cells";
RL Kidney Int. 46:1307-1314(1994).
RN [10]
RP REVIEW ON VARIANTS.
RX MEDLINE=97338652; PubMed=9195222;
RA DOI=10.1002/(SICI)1098-1004(1997)9:6<477::AID-HUMU13.3.CO;2-H;
RA Lemmink H.H., Schroeder C.H., Monnens L.A.H., Smeets H.J.M.;
RT "The clinical spectrum of type IV collagen mutations";
RL Hum. Mutat. 9:477-499(1997).
RN [11]
RP VARIANT AS SER-1564.
RX MEDLINE=91169492; PubMed=1672282;
RA Zhou J., Barker D.F., Hostikka S.L., Gregory M.C., Atkin C.L.,
RA Tryggvason K.;
RT "Single base mutation in alpha 5(IV) collagen chain gene converting a
RT conserved cysteine to serine in Alport syndrome";
RL Genomics 9:10-18(1991).
RN [12]
RP VARIANT AS ARG-325.
RX MEDLINE=92303559; PubMed=1376965;
RA Knebelmann B., Deschenes G., Gros F., Hore M.-C., Gruenfeld J.-P.,
RA Tryggvason K., Gubler M.-C., Aulagnac C.;
RT "Substitution of arginine for glycine 325 in the collagen alpha 5 (IV)
chain associated with X-linked Alport syndrome: characterization of
RT the mutation by direct sequencing of PCR-amplified lymphoblast cDNA
fragments";
RL Am. J. Hum. Genet. 51:135-142(1992).
RN [13]
RP VARIANT AS GLU-325.
RX MEDLINE=93244772; PubMed=1363780;
RA Renieri A., Serti M., Myers J.C., Pihlajaniemi T., Maseella L.,
RA Rizzoni G.F., de Marchi M.;
RT "De novo mutation in the COL4A5 gene converting glycine 325 to
RT glutamic acid in Alport syndrome";
RL Hum. Mol. Genet. 1:127-129(1992).
RN [14]
RP VARIANTS AS THR-1517, SER-1538 AND GLN-1563.
RX MEDLINE=94010948; PubMed=8406498;
RA Lemmink H.L., Schroeder C.H., Brunner H.G., Nelen M.R., Zhou J.,
RA Tryggvason K., Haggama-Schouten W.A.G., Roodvoets A.P., Rascher W.,
RA van Oost B.A., Smeets H.J.M.;
RT "Identification of four novel mutations in the COL4A5 gene of patients
RT with Alport syndrome";
RL Genomics 17:485-489(1993).
RN [15]

RP VARIANTS AS GLU-400; VAL-406; VAL-638; ALA-638; ARG-653; ARG-796;
 RP ARG-869; ARG-872 AND CYS-1241.
 RX MEDLINE=95322976; PubMed=7596631;
 RA Boye E., Flinter F., Zhou J., Trygsvaen K., Bobrow M., Harris A.;
 RT "Detection of 12 novel mutations in the collagenous domain of the
 RL COL4A5 gene in Alport syndrome patients."; Hum. Mutat. 5:197-204(1995).
 RN [16]
 RP VARIANT AS ARG-1649.
 RX MEDLINE=966213750; PubMed=8651292;
 RA Barber D.P., Pruchno C.J., Jiang X., Atkin C.L., Stone E.M.,
 RT "A mutation causing Alport syndrome with tardive hearing loss is
 RL common in the western United States."; Am. J. Hum. Genet. 58:1157-1165(1996).
 RN [17]
 RP VARIANTS AS
 RX MEDLINE=96213754; PubMed=8651296;
 RA Renieri A., Brutini M., Galli L., Zanelli P., Neri T.M., Rossetti S.,
 RX Turco A.E., Heiskari N., Zhou J., Guemano R., Massella L., Banfi G.,
 RA Scioari F., Sessa A., Rizzoni G.F., Trygsvaen K., Pignatelli P.F.,
 RA Savi M., Ballabio A., de Marchi M.;
 RT "X-linked Alport syndrome: an SSCP-based mutation survey over all 51
 RL exons of the COL4A5 gene."; Am. J. Hum. Genet. 58:1192-1204(1996).
 RN [18]
 RP VARIANTS AS, AND VARIANTS ASP-430; SER-444; SER-619; ASN-664 AND
 RP MET-1428.
 RX MEDLINE=97094179; PubMed=8940267;
 RA Kneelmann B., Breillat C., Forestier L., Arrondel C., Jaccardier D.,
 RA Glazias I., Drouot L., Deschenes G., Gruenfeld J.-P., Broyer M.,
 RA Gubler M.-C., Antignac C.;
 RT "Spectrum of mutations in the COL4A5 collagen gene in X-linked Alport
 RL syndrome."; Am. J. Hum. Genet. 59:1221-1232(1996).
 RN [19]
 RP VARIANT AS ASP-1498.
 RX MEDLINE=96233932; PubMed=8829632;
 RL DOI=10.1002/(SICI)1098-1004(1996)7:2<149::AID-HUMU9>3.3.CO;2-A;
 RA Tverskaya S., Bobrykina V., Tsalykova F., Ignatova M.,
 RA Krasnopolskaya X., Bygratov O.;
 RT "Substitution of A1498D in noncollagen domain of $\alpha 5(\text{IV})$ collagen chain
 RL associated with adult-onset X-linked Alport syndrome."; Hum. Mutat. 7:149-150(1996).
 RN [20]
 RP VARIANT AS GLN-1677.
 RX MEDLINE=97295089; PubMed=9150741; DOI=10.1007/s004390050429;
 RA Barter D.F., Denison J.C., Atkin C.L., Gregory M.C.;
 RT "Common ancestry of three Ashkenazi-American families with Alport
 RL syndrome and COL4A5 R1677Q."; Hum. Genet. 99:681-684(1997).
 RN [21]
 RP VARIANTS AS ARG-174; ARG-177; ARG-325; CYS-1410; TRP-1421; THR-1517
 RP AND ASP-1596.
 RX MEDLINE=98112435; PubMed=9452056;
 RA Neri T.M., Zanelli P., de Palma G., Savi M., Rossetti S., Turco A.E.,
 RA Pignatelli G.F., Galli L., Brutini M., Renieri A., Mingarelli R.,
 RA Trivelli A., Pinciaroli A.R., Ragatolo M., Rizzoni G.F., de Marchi M.;
 RT "Missense mutations in the COL4A5 gene in patients with X-linked
 RL Alport syndrome."; Hum. Mutat. Suppl. 1:S106-S109(1998).
 RN [22]
 RP VARIANTS AS VAL-420; 456-PRO-PRO-458 DEL; ASP-573; ASP-624; ASP-635;
 RP 802-GLY-PRO-807 DEL; ARG-869; CYS-941; SER-1030; SER-1066; ASP-1143;
 RP ARG-1196; GLU-1261; SER-1357 AND ARG-1649.
 RX MEDLINE=99063529; PubMed=9848783;
 RA Martin P., Heiskari N., Zhou J., Leinonen A., Tanelius T., Hertz J.M.,
 RA Barker D.F., Gregory M.C., Atkin C.L., Stykardottir U., Neumann H.,
 RA Sprigade J., Shows T.B., Petersen E., Trygsvaen K.;
 RT "High mutation detection rate in the COL4A5 collagen gene in suspected
 RL Alport syndrome using PCR and direct DNA sequencing."; J. Am. Soc. Nephrol. 9:2291-2301(1998).

Query Match 41.0%; Score 390; DB 1; Length 1685;
 Best Local Similarity 43.2%; Pred. No. 2.2e-15;
 Matches 80; Conservative 11; Mismatches 40; Indels 54; Gaps 4;
 QY 2 GPPGKGTGTHPG-----LPPKGDCKPGRPSTGTGPAEGEPGAMPGQGRPPGPVH 55
 DB 147 GPPGIPGKKGKGPBSIMSLPFGPKNPFYPPGPGIGLPGRTGIPGPGF---PPPLM 203
 QY 56 GPPGPPGPPGAGISAVGLKDRGATGERGLAP----- 90
 DB 204 GPPGPPGLPGPKGNMGNLFQGPKEKGEGLQGPPEPGQISEQKRPIDVEFQKGDQGLP 263
 QY 91 ---GPPGPPGPPGPG-----YKMGATGPMGQGGIRGIRGPPGPMGQ 130
 DB 264 GDRGPPGPPGIRGPPGPGKEKGEKGEQGPCKRGKPKDGNQGIPLGPDYFGE 323
 QY 131 PGKAG 135
 DB 324 PGKDG 328
 RESULT 13
 Q9NUB7 HUMAN
 ID Q9NUB7 HUMAN PRELIMINARY; PRT; 1685 AA.
 AC Q9NUB7_Q72700; DB 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TrEMBLrel. 31, Last annotation update)
 DE 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
 DE Collagen, type IV, alpha 5 (Alport syndrome).
 GN Name=COL4A5; ORFNames=RP6-24A23.5-001;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominiidae;
 OC Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Cobley V.;
 RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RA Bird C.;
 RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL035425; CAB90289.2; -; Genomic DNA.
 DR EMBL; AL034369; CA142267.2; -; Genomic DNA.
 DR EMBL; AL031622; CA143038.1; -; Genomic DNA.
 DR EMBL; AL034369; CAB90289.2; JOINED; Genomic DNA.
 DR EMBL; AL031622; CAB90289.2; JOINED; Genomic DNA.
 DR EMBL; AL031622; CA142267.2; JOINED; Genomic DNA.
 DR EMBL; AL034369; CA142267.2; JOINED; Genomic DNA.
 DR EMBL; AL034369; CA143038.1; JOINED; Genomic DNA.
 DR EMBL; AL034369; CA143038.1; JOINED; Genomic DNA.
 DR SMR; Q9NUB7; 1458-1685.
 DR Ensemble; ENSG00000188153; Homo sapiens.
 DR GO; GO:0005581; C:collagen; IEA.
 DR GO; GO:0005737; C:cytoplasm; IEA.
 DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
 DR GO; GO:0006817; P:phosphate transport; IEA.
 DR InterPro; IPR008161; C1g helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR001442; Procollagen4_C.
 DR Pfam; PF01413; C4; 2.
 DR Pfam; PF01391; Collagen; 22.
 DR ProDom; PDO00007; C1g helix; 3.
 DR ProDom; PDO003923; Procollagen4_C; 2.
 DR SMART; SM00111; C4; 2.
 KW Collagen.
 SQ SEQUENCE 1685 AA; 161044 MW; 4450A6762F12A626 CRC64;
 Query Match 41.0%; Score 390; DB 2; Length 1685;
 Best Local Similarity 43.2%; Pred. No. 2.2e-15;
 Matches 80; Conservative 11; Mismatches 40; Indels 54; Gaps 4;

```

QY 2 GPPGFKGTGHPG-----LPGPKGDCGKPPPGSTGRPGAGEBPGAMGPGPPGHV 55
DB 147 GPPGIGMGKGPESISIMSLPGPKGNBVGYPGPGIGQLPPTGIPGPIGP-----PGPPGLM 203
QY 56 GPPGPPGPGPPGPGISAVGLKGDGATGERTGLP-----90
DB 204 GPPGPPGGLPBPKNMGLNFGPKGKEBGLGPPGPPGQISQKPRIDVFGKGDGLP 263
QY 91 ---GPPGPPGPPGPPG-----YKMKATGPMGQGGIPGIPGPPGPMGQ 130
DB 264 GDDGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 323
QY 131 PGKAG 135
DB 324 PGKAG 328

RESULT 14
COH1_CHICK STANDARD; PRT; 1146 AA.
AC 090584;
DT 29-MAR-2004 (Rel. 43, Last sequence update)
DT 29-MAR-2004 (Rel. 43, Last sequence update)
DE 10-MAY-2005 (Rel. 47, Last annotation update)
DE Collagen alpha 1(XVII) chain (Bullous pemphigoid antigen 2) (180 kDa
DE bullous pemphigoid antigen 2) (Fragment).
GN Name=COL17A1; Synonym=BP180, BPAG2;
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP NUCLEOTIDE SEQUENCE, AND TISSUE SPECIFICITY.
RC TISSUE=Cornea;
RX MEDLINE=91142213; PubMed=1705041;
RA Marchant J.K., Linsemayer T.F., Gordon M.K.;
RT "CDNA analysis predicts a cornea-specific collagen.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:1560-1564 (1991).
CC -1- FUNCTION: May play a role in the integrity of hemidesmosome and
CC the attachment of basal keratinocytes to the underlying basement
CC membrane (By similarity).
CC -1- SUBUNIT: Homotrimer of alpha 1(XVII) chains (By similarity).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (By similarity).
CC -1- TISSUE SPECIFICITY: Cornea specific.
CC -1- PTM: The intracellular/endo domain is disulfide-linked (By
CC similarity).
CC -1- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -1- PTM: Undergoes proteolytic cleavage by furin-like protease to
CC yield a 120 kDa soluble form that corresponds to the ectodomain
CC (By similarity).
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
CC EMBL; M60172; AAA48703.1; -; mRNA.
DR PIR; S16501; A38587.
DR Ensembl; ENSGALG0000008323; Gallus gallus.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR Pfam; PF01391; Collagen_6.
DR Prodom; PD000007; C1g_helix; 3.
KW Collagen; Extracellular matrix; Glycoprotein; Hydroxylation; Repeat;
KW Signal-anchor; Structural protein; Transmembrane.
FT TOPO_DOM <1 52 Cytoplasmic (Potential).
FT TRANSMEM 53 74 Signal-anchor for type II membrane
FT PROTEIN 1 146 Extracellular (Potential).
FT REGION 1 149 Nonhelical region (NC16).

```

```

FT REGION 150 1101 Triple-helical region.
FT REGION 1102 1146 Nonhelical region (NC1).
FT CARBOHYD 1041 1041 N-linked (GlcNAc...) (Potential).
FT NON_TER 1 1
SQ SEQUENCE 1146 AA; 114814 MW; 1E0E9977D537A859 CRC64;

Query Match 40.7%; Score 387; DB 1; Length 1146;
Best Local Similarity 40.1%; Pred. No. 2,4e-15;
Matches 89; Conservative 10; Mismatches 57; Indels 66; Gaps 7;

QY 2 GPPGFKGTGHPGLPBPCKDCGKPPPGSTGRPGAGEBPGAMPP-----QGRP 49
DB 350 GPPGPGDGTGEBPLTGPQ---GPPGLPGNPRGAGABGAPGVASAGSSTIALPGP 406
QY 50 GPPGHPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 88
DB 407 GPPGPIGPTGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 464
QY 89 LPQPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 127
DB 465 LQGRAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 524
QY 128 MGQPGKAGHCNPSDCTGAMPMEQIYPMK-----TMKPPFG 163
DB 525 PGPKGDPGPPGPPG-FTGEPFEGGLPGFSSHGCTVTMQPPG 565

RESULT 15
COH1_CHICK STANDARD; PRT; 1453 AA.
AC P02457;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Collagen alpha 1(I) chain precursor.
GN Name=COL1A1;
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP NUCLEOTIDE SEQUENCE OF 1-153.
RX MEDLINE=88056316; PubMed=3678834; DOI=10.1016/0378-1119(87)90159-4;
RA Finer M.H., Boedtker H., Doty P.;
RT "Unusual DNA sequences located within the promoter region and the
RT first intron of the chicken pro-alpha 1(I) collagen gene.";
RL J. Biol. Chem. 262:13323-13332 (1987).
RN [2]
RP NUCLEOTIDE SEQUENCE OF 1-144.
RX MEDLINE=88007542; PubMed=2820966;
RA Finer M.H., Aho S., Gerstenfeld L.C., Boedtker H., Doty P.;
RT "Unusual DNA sequences located within the promoter region and the
RT first intron of the chicken pro-alpha 1(I) collagen gene.";
RL J. Biol. Chem. 262:13323-13332 (1987).
RN [3]
RP PROTEIN SEQUENCE OF 152-1187.
RX MEDLINE=82231995; PubMed=7093229;
RA Higberger J.H., Corbett C., Dixit S.N., Yu W., Seyer J.M., Kang A.H.,
RA Gross J.;
RT "Amino acid sequence of chick skin collagen alpha 1(I)-C88 and the
RT complete primary structure of the helical portion of the chick skin
RT collagen alpha 1(I) chain.";
RL Biochemistry 21:2048-2055 (1982).
RN [4]
RP PROTEIN SEQUENCE OF 1200-1205.
RX MEDLINE=72243016; PubMed=5047697;
RA Eyre D.R., Glimcher M.J.;
RT "Evidence for a previously undetected sequence at the carboxyterminus
RT of the alpha 1 chain of chicken bone collagen.";
RL Biochem. Biophys. Res. Commun. 48:720-726 (1972).
RN [5]
RP NUCLEOTIDE SEQUENCE OF 981-1453.

```

```

RX MEDLINE=81160715; PubMed=6927845;
RA Fuller F., Boedtker H.;
RT "Sequence determination and analysis of the 3' region of chicken pro-
RT alpha 1(I) and pro-alpha 2(I) collagen messenger ribonucleic acids
RT including the carboxy-terminal propeptide sequences."
RL Biochemistry 20:996-1006(1981).
RN [6]
RP NUCLEOTIDE SEQUENCE OF 1311-1453.
RX MEDLINE=80134546; PubMed=6987088; DOI=10.1016/0014-5793(80)80761-7;
RA Showalter A.M., Pesciotta D.M., Eikenberry E.F., Yamamoto T.,
RA Paetan I., Derombrughe B., Pietrek P.P., Olsen B.R.;
RT "Nucleotide sequence of a collagen cDNA-fragment coding for the
RT carboxyl end of pro alpha 1(I)-chains."
RL FEBS Lett. 111:61-65(1980).
CC -!- FUNCTION: Type I collagen is a member of group I collagen
CC (fibrillar forming collagen).
CC -!- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.
CC -!- TISSUE SPECIFICITY: Forms the fibrils of tendon, ligaments and
CC bones. In bones the fibrils are mineralized with calcium
CC hydroxyapatite.
CC -!- PTM: Proline residues at the third position of the tripeptide
CC repeating unit (G-X-Y) are hydroxylated in some or all of the
CC chains. Pro-1153 is the only 3-hydroxypro and the only
CC hydroxylated proline in position X.
CC -!- SIMILARITY: Belongs to the fibrillar collagen family.
CC -!- SIMILARITY: Contains 1 WMP domain.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
DR EMBL: M17839; AAA48704.1; -; Genomic DNA.
DR EMBL: M17838; AAA48704.1; JOINED; Genomic_DNA.
DR EMBL: V00401; CAA23695.1; -; mRNA.
DR EMBL: M10571; AAA48671.1; ALT_SEQ; mRNA.
DR EMBL: M17607; AAA48672.1; -; mRNA.
DR PIR: A27179; A27179.
DR PIR: I50629; I50629.
DR InterPro: IPR008161; C1g_helix.
DR InterPro: IPR008160; Collagen.
DR InterPro: IPR000885; Fib_collagen_C.
DR InterPro: IPR001007; WMP_C.
DR Pfam: PF01410; COLFI; 1.
DR Pfam: PF01391; Collagen; 18.
DR Pfam: PF00093; WMP; 1.
DR ProDom: PD000007; C1g_helix; 2.
DR ProDom: PD002078; Fib_collagen_C; 1.
DR SMART; SM00214; COLFI; 1.
DR SMART; PS01208; WMP; 1.
DR PROSITE; PS0184; WMP; 2; 1.
DR Collagen; Direct protein sequencing; Extracellular matrix;
KW Glycoprotein; Hydroxylation; Pyrolydine carboxylic acid; Repeat;
KW Signal; Structural protein.
FT SIGNAL 1 22
FT PROPEP 23 151 N-terminal propeptide.
FT CHAIN 152 1205 Collagen alpha 1(I) chain.
FT PROPEP 1206 1453 C-terminal propeptide.
FT DOMAIN 31 89 WMP.
FT MOD_RES 152 152 Pyrolydine carboxylic acid.
FT MOD_RES 160 160 Alysine (By similarity).
FT MOD_RES 254 254 5-hydroxylysine (By similarity).
FT MOD_RES 254 254 5-hydroxylysine (Potential).
FT MOD_RES 851 851 5-hydroxyproline (Potential).
FT MOD_RES 1081 1081 Hydroxyproline (Potential).
FT MOD_RES 1097 1097 5-hydroxylysine (Potential).
FT MOD_RES 1153 1153 3-hydroxyproline.
FT MOD_RES 1153 1153 O-linked (Gal.. ) (By similarity).
FT CARBOHYD 254 254 N-linked (GlcNAc.. ) (By similarity).
FT CARBOHYD 1354 1354 F -> L (in Ref. 5).
FT CONFLICT 1187 1187 F -> H (in Ref. 6).
FT CONFLICT 1441 1441 Q -> H (in Ref. 6).
SQ SEQUENCE 1453 AA; 137789 MW; 3BC6152134271F4D CRC64;

```

```

Query Match 40.7%; Score 387; DB 1; Length 1453;
Best Local Similarity 46.9%; Pred. No. 2.9e-15;
Matches 82; Conservative 12; Mismatches 59; Indels 22; Gaps 5;

QY 2 GPP---GPKGTHGHLPGPKDCGKPPPGSTGRPGAEGEPGAMGPGRPGPHVGP 58
DB 858 GPPGATGPPGAAGVGPFGPSGNIGLPPPPACKZSKGRGRTGPPAGRGPPAGPP 917
QY 59 GPPGQPGPAG----ISAVGLKDRGANGERGLAGLPGQ-----PGPPGQPGPGYGMG 108
DB 918 GPPGEGKSPGADGPIGAPGTGPGQIAGRGVGLPGQRGGRGFPGLPGPSGEP--GKG 975
QY 109 ATGPMGQCGIRGIRGPPPPMGQPKAGHCNPSDCGAMPMPGQYPPMKTMKGPPG 163
DB 976 PSGASGERGPPGMPGPIAGPPGEGREGAPGAEGA-----PRDGAAGPKG 1023

Search completed: December 1, 2005, 08:26:47
Job time : 233 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Comugen Ltd.

OM protein - protein search, using sw model

Run on: December 1, 2005, 08:28:25 ; Search time 188 seconds
(without alignments)
380.951 Million cell updates/sec

Title: US-09-924-340-58

Sequence: 1 MGPPGFKGTGHPGLPGPKG.....GAMPMEQYPPMKTMGPRG 163

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 2443163 seqs, 439378781 residues

Word size : 100

Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database :

A_Geneseq_21:*

- 1: geneseqp1980s:*
- 2: geneseqp1980s:*
- 3: geneseqp2000s:*
- 4: geneseqp2001s:*
- 5: geneseqp2002s:*
- 6: geneseqp2003as:*
- 7: geneseqp2003bs:*
- 8: geneseqp2004s:*
- 9: geneseqp2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	163	100.0	163	ABR48481	ABR48481 Human Alp

ALIGNMENTS

RESULT 1
ABR48481
ID ABR48481 standard; protein; 163 AA.
XX
AC ABR48481;
XX
DT 13-JUN-2003 (first entry)
XX
DE Human Alpha 1 type XVI collagen.
XX
KW Human; GENSET; therapeutic; therapy.
XX
OS Homo sapiens.
XX
PN WO200294864-A2.
XX
PD 28-NOV-2002.
XX

PF 06-AUG-2001; 2001WO-1B001715.
XX
PR 25-MAY-2001; 2001US-0293574P.
PR 15-JUN-2001; 2001US-0298698P.
PR 29-JUN-2001; 2001US-0302277P.
PR 13-JUL-2001; 2001US-0305456P.
XX
PA (GEST) GENSET.
XX
PI Benjamin S, Tanaka H;
XX
DR MPI; 2003-129412/12.
DR N-PDB; ACC51088.
XX

PT New GENSET polynucleotides and polypeptides, useful for preparing a composition for treating GENSET-related disorders and as reagents in assays to quantitatively determined levels of GENSET expression in biological samples.
PT
PS Claim 2; Page 451; 505pp; English.
XX

CC The present invention relates to novel human GENSET coding sequences (ACC51060-ACC51115) and proteins (ABR48453-ABR48508). The GENSET sequences are useful for preparing a composition for treating GENSET-related disorders. They can also be used as markers for tissues in which the corresponding protein is preferentially expressed, as molecular weight markers on Southern gels, as chromosome markers or tags to identify chromosomes, and as reagents in assays to quantitatively determined levels of GENSET expression in biological samples
CC
XX
SQ Sequence 163 AA:

Query Match 100.0%; Score 163; DB 6; Length 163;
Best Local Similarity 100.0%; Pred.No.5,2e-136;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MGPPGFKGTGHPGLPGPKDCKPDPGSTRGAGGERGAMGPOGRPGHVGPPGP	60
DB	1	MGPPGFKGTGHPGLPGPKDCKPDPGSTRGAGGERGAMGPOGRPGHVGPPGP	60
QY	61	PGQPGAGISAVGLKCDRGATGERSGLAGLPQGPDPGPPGQPGPYGKMGATGPMGQGIPIG	120
DB	61	PGQPGAGISAVGLKCDRGATGERSGLAGLPQGPDPGPPGQPGPYGKMGATGPMGQGIPIG	120
QY	121	IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPRG	163
DB	121	IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPRG	163

Search completed: December 1, 2005, 08:48:14
Job time : 188 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 1, 2005, 08:38:15 ; Search time 46 Seconds
(without alignments)
292.959 Million cell updates/sec

Title: US-09-924-340-58

Perfect score: 163
Sequence: 1 MGPPGFKGTGHPGLPGPKG.....GAMPMEQYPPMKTMKPGF 163

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 572060 seqs, 82675679 residues

Word size : 100

Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database : Issued Patents NA:*

- 1: /cgn2_6/prodata/1/1aa/5 COMB.pep.*
- 2: /cgn2_6/prodata/1/1aa/6 COMB.pep.*
- 3: /cgn2_6/prodata/1/1aa/7 COMB.pep.*
- 4: /cgn2_6/prodata/1/1aa/8 COMB.pep.*
- 5: /cgn2_6/prodata/1/1aa/9 COMB.pep.*
- 6: /cgn2_6/prodata/1/1aa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	163	100.0	163	2	US-10-000-489-58 Sequence 58, Appl

ALIGNMENTS

RESULT 1
US-10-000-489-58
; Sequence 58, Application US/10000489
; Patent No. 6794363
; GENERAL INFORMATION:
; APPLICANT: Benjamin, Stephanie
; APPLICANT: Tanaka, Hiroaki
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
; FILE REFERENCE: 91.US6.DIV
; CURRENT APPLICATION NUMBER: US/10/000,489
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: US 09/924,340
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: PCT/IB01/01715
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/305,456
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/302,277
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/298,698
; PRIOR FILING DATE: 2001-06-15

; PRIOR APPLICATION NUMBER: US 60/293,574
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: JPatent
; SEQ ID NO 58
; LENGTH: 163
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-000-489-58

Query Match 100.0%; Score 163; DB 2; Length 163;
Best local similarity 100.0%; Pred. No. 5.2e-138;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MGPPGFKGTGHPGLPGPKDCGKPPGSTGRPGAEGEPGAMGPGQRPGRHVGPPG 60
DB	1	MGPPGFKGTGHPGLPGPKDCGKPPGSTGRPGAEGEPGAMGPGQRPGRHVGPPG 60
QY	61	PGQPPAGISAVGLKGDGRGATGSRGLAGLPQGPQPPQGPQPGYGMKATGPMGQGSIFG 120
DB	61	PGQPPAGISAVGLKGDGRGATGSRGLAGLPQGPQPPQGPQPGYGMKATGPMGQGSIFG 120
QY	121	IPGPPGPMQGPGRAGHCNPSDCFGAMPMEQYPPMKTMKPGF 163
DB	121	IPGPPGPMQGPGRAGHCNPSDCFGAMPMEQYPPMKTMKPGF 163

Search completed: December 1, 2005, 08:49:50
Job time : 47 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OW protein - protein search, using sw model

Run on: December 1, 2005, 08:48:25 ; Search time 165 Seconds
(without alignments)
412.765 Million cell updates/sec

Title: US-09-924-340-58
Perfect score: 163
Sequence: 1 MGPPGFKGTGHPGLPGPKG.....GAMPMEQYPPMKTKGPF 163

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1867569 seqs, 417829326 residues

Word size : 100

Total number of hits satisfying chosen parameters: 8

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database :

Published Applications AA Main:*

- 1: /cgn2_6/prodata/1/pubpaa/us07_PUBCOMB.pep:*
- 2: /cgn2_6/prodata/1/pubpaa/us08_PUBCOMB.pep:*
- 3: /cgn2_6/prodata/1/pubpaa/us09_PUBCOMB.pep:*
- 4: /cgn2_6/prodata/1/pubpaa/us10_PUBCOMB.pep:*
- 5: /cgn2_6/prodata/1/pubpaa/us10B_PUBCOMB.pep:*
- 6: /cgn2_6/prodata/1/pubpaa/us11_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	163	100.0	163	3	US-09-924-340-58
2	163	100.0	163	3	US-09-924-340-58
3	163	100.0	163	3	US-09-924-340-58
4	163	100.0	163	3	US-09-924-340-58
5	163	100.0	163	3	US-09-924-340-58
6	163	100.0	163	3	US-09-924-340-58
7	163	100.0	163	3	US-09-924-340-58
8	163	100.0	163	3	US-09-924-340-58

ALIGNMENTS

RESULT 1
US-09-924-340-58
Sequence 58, Application US/09992600A
Publication No. US20030027161A1
GENERAL INFORMATION:
APPLICANT: Benjamin, Stephanie
APPLICANT: Tanaka, Hiroaki
TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
FILE REFERENCE: 91.US4.DIV
CURRENT APPLICATION NUMBER: US/09/924,600A
CURRENT FILING DATE: 2001-11-13
PRIOR APPLICATION NUMBER: US 09/924,340
PRIOR FILING DATE: 2001-08-06
PRIOR APPLICATION NUMBER: PCT/IB01/01715

PRIOR FILING DATE: 2001-08-06
PRIOR APPLICATION NUMBER: US 60/305,456
PRIOR FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: US 60/302,277
PRIOR FILING DATE: 2001-06-29
PRIOR APPLICATION NUMBER: US 60/298,698
PRIOR FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: US 60/293,574
NUMBER OF SEQ ID NOS: 114
SOFTWARE: JPatent
SEQ ID NO 58
LENGTH: 163
TYPE: PRT
ORGANISM: Homo sapiens
US-09-924-340-58

Query Match 100.0%; Score 163; DB 3; Length 163;
Best Local Similarity 100.0%; Pred. No. 1.2e-136;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGPPGFKGTGHPGLPGPKDCKPSTGRPGAEGEPGAMPQGPGRPGPHVGP 60
DB 1 MGPPGFKGTGHPGLPGPKDCKPSTGRPGAEGEPGAMPQGPGRPGPHVGP 60
QY 61 PGQPGAGISAVGLKGDGRGATGERGLAGLPGQGPDPGPGPGPGYGMKATGPMGQGSIFG 120
DB 61 PGQPGAGISAVGLKGDGRGATGERGLAGLPGQGPDPGPGPGPGYGMKATGPMGQGSIFG 120
QY 121 IPQPGPMQPGKAGHCNPSDCFGAMPMEQYPPMKTKGPF 163
DB 121 IPQPGPMQPGKAGHCNPSDCFGAMPMEQYPPMKTKGPF 163

RESULT 2
US-09-924-340-58
Sequence 58, Application US/09924340
Publication No. US20030027248A1
GENERAL INFORMATION:
APPLICANT: Benjamin, Stephanie
APPLICANT: Tanaka, Hiroaki
TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
FILE REFERENCE: 91.US2.REG
CURRENT APPLICATION NUMBER: US/09/924,340
CURRENT FILING DATE: 2001-08-06
PRIOR APPLICATION NUMBER: US 60/305,456
PRIOR FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: US 60/302,277
PRIOR FILING DATE: 2001-06-29
PRIOR APPLICATION NUMBER: US 60/298,698
PRIOR FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: US 60/293,574
NUMBER OF SEQ ID NOS: 112
SOFTWARE: JPatent
SEQ ID NO 58
LENGTH: 163
TYPE: PRT
ORGANISM: Homo sapiens
US-09-924-340-58

Query Match 100.0%; Score 163; DB 3; Length 163;
Best Local Similarity 100.0%; Pred. No. 1.2e-136;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MGPPGFKGTGHPGLPGPKDCKPSTGRPGAEGEPGAMPQGPGRPGPHVGP 60
DB 1 MGPPGFKGTGHPGLPGPKDCKPSTGRPGAEGEPGAMPQGPGRPGPHVGP 60
QY 61 PGQPGAGISAVGLKGDGRGATGERGLAGLPGQGPDPGPGPGPGYGMKATGPMGQGSIFG 120
DB 61 PGQPGAGISAVGLKGDGRGATGERGLAGLPGQGPDPGPGPGPGYGMKATGPMGQGSIFG 120

QY 121 I PGP PGP MG D P G K A G H C N P S D C F G A M P M E Q Y P P M K T M K G P F G 163

Db 121 I P G P P G P M G Q P G K A G H C N P S D C F G A M P M E Q Y P P M K T M K G P F G 163

RESULT 3

US-09-992-095B-58
; Sequence 58, Application US/09992095B
; Publication No. US20030157485A1

```

: NUMBER OF SEQ ID NOS: 112
: SOFTWARE: Jpatent
: SEQ ID NO: 58
: LENGTH: 163
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-09-992-095B-58

```

RESULT 4
US-09-999-570-58

APPLICANT: Tanaka, Hiroaki
 TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
 FILE REFERENCE: G-0919US08019
 CURRENT APPLICATION NUMBER: US/09/999,570
 CURRENT FILING DATE: 2001-06-14
 PRIOR APPLICATION NUMBER: US 09/924,340
 PRIOR FILING DATE: 2001-08-06
 PRIOR APPLICATION NUMBER: PCT/IB01/01715
 PRIOR FILING DATE: 2001-08-06
 PRIOR APPLICATION NUMBER: US 60/305,456
 PRIOR FILING DATE: 2001-07-13
 PRIOR APPLICATION NUMBER: US 60/302,277
 PRIOR FILING DATE: 2001-06-29
 PRIOR APPLICATION NUMBER: US 60/298,698

```

; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/293,574
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 112

```

ORGANISM: Homo sapiens
US-09-999-570-58

Qy 121 I P G P P G M G P G K A G H C N P S D C F G A M P M E Q Q Y P P M K T M K G P F G 163

Db 121 I P G P P G M G P G K A G H C N P S D C F G A M P M E Q Q Y P P M K T M K G P F G 163

US-10-000-489-58

	Query Match	Score	DB 4	Length	163					
	Best Local Similarity	100.0%	Pred. No. 1.2e-136							
	Matches	163	Conservative	0	Mismatches	0	Indels	0	Gaps	0
Qy	1	MGPFGFKGKTCHPELPGPKDKDCKRPPPGSGTGRGAESEBPAMGPQGRPPRGHNGPGR	60							
Db	1	MGPFGFKGKTCHPELPGPKDKDCKRPPPGSGTGRGAESEBPAMGPQGRPPRGHNGPGR	60							
Qy	61	PGQGPAGISAVLKGKGRGATGERGLAGLPGQPPPGQGPYGKMGATPPMGQGGIPG	120							
Db	61	PGQGPAGISAVLKGKGRGATGERGLAGLPGQPPPGQGPYGKMGATPPMGQGGIPG	120							
Qy	121	IPGPGPMGQPGKAGHCNPSCFCGAMPMEBOYPPMKTIWKGDFG	163							

Db 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQOYPPMKTMGPF 163

RESULT 6
US-10-000-986-58
Sequence 58, Application US/10000986

GENERAL INFORMATION:
Publication No. US20030096247A1
APPLICANT: Benjamin, Hiroaki
TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
FILE REFERENCE: 91.US9.DIV
CURRENT APPLICATION NUMBER: US/10/000,986
CURRENT FILING DATE: 2001-11-14
PRIOR APPLICATION NUMBER: US 09/924,340
PRIOR FILING DATE: 2001-08-06
PRIOR APPLICATION NUMBER: PCT/IB01/01715
PRIOR FILING DATE: 2001-08-06
PRIOR APPLICATION NUMBER: US 60/305,456
PRIOR FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: US 60/302,277
PRIOR FILING DATE: 2001-06-29
PRIOR APPLICATION NUMBER: US 60/298,698
PRIOR FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: US 60/293,574
PRIOR FILING DATE: 2001-05-25
NUMBER OF SEQ ID NOS: 112
SOFTWARE: JPatent
SEQ ID NO 58
LENGTH: 163
TYPE: PRT
ORGANISM: Homo sapiens
US-10-000-986-58

Query Match 100.0%; Score 163; DB 4; Length 163;
Best Local Similarity 100.0%; Pred. No. 1.2e-136;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGPPGKGTGHTGHLGPPKDCGKPPPGSTGRPGAEGEPGAMGPGCRPGPHVGP 60
Db 1 MGPPGKGTGHTGHLGPPKDCGKPPPGSTGRPGAEGEPGAMGPGCRPGPHVGP 60
Qy 61 PCQPGAGISAVGLKGDRTGATGSRGLAGLPGQGPFGPGPGPGYGTGKATGTPWGQGGIPG 120
Db 61 PCQPGAGISAVGLKGDRTGATGSRGLAGLPGQGPFGPGPGPGYGTGKATGTPWGQGGIPG 120
Qy 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQOYPPMKTMGPF 163
Db 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQOYPPMKTMGPF 163

RESULT 7
US-10-154-678-58
Sequence 58, Application US/10154678

GENERAL INFORMATION:
Publication No. US20030162186A1
APPLICANT: Benjamin, Stephanie
TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
FILE REFERENCE: 182.US1.REG
CURRENT APPLICATION NUMBER: US/10/154,678
CURRENT FILING DATE: 2002-10-15
PRIOR APPLICATION NUMBER: US 09/924,340
PRIOR FILING DATE: 2001-08-06
PRIOR APPLICATION NUMBER: US 60/305,456
PRIOR FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: US 60/302,277
PRIOR FILING DATE: 2001-06-29
PRIOR APPLICATION NUMBER: US 60/298,698
PRIOR FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: US 60/293,574
PRIOR FILING DATE: 2001-05-25
NUMBER OF SEQ ID NOS: 112

SOFTWARE: JPatent
SEQ ID NO 58
LENGTH: 163
TYPE: PRT
ORGANISM: Homo sapiens
US-10-154-678-58

Query Match 100.0%; Score 163; DB 4; Length 163;
Best Local Similarity 100.0%; Pred. No. 1.2e-136;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGPPGKGTGHTGHLGPPKDCGKPPPGSTGRPGAEGEPGAMGPGCRPGPHVGP 60
Db 1 MGPPGKGTGHTGHLGPPKDCGKPPPGSTGRPGAEGEPGAMGPGCRPGPHVGP 60
Qy 61 PCQPGAGISAVGLKGDRTGATGSRGLAGLPGQGPFGPGPGPGYGTGKATGTPWGQGGIPG 120
Db 61 PCQPGAGISAVGLKGDRTGATGSRGLAGLPGQGPFGPGPGPGYGTGKATGTPWGQGGIPG 120
Qy 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQOYPPMKTMGPF 163
Db 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQOYPPMKTMGPF 163

RESULT 8
US-10-838-854-58
Sequence 58, Application US/10838854

GENERAL INFORMATION:
Publication No. US20050026182A1
APPLICANT: Benjamin, Stephanie
TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
FILE REFERENCE: 91.US6.DIV
CURRENT APPLICATION NUMBER: US/10/838,854
CURRENT FILING DATE: 2004-05-03
PRIOR APPLICATION NUMBER: US/10/000,489
PRIOR FILING DATE: 2001-11-14
PRIOR APPLICATION NUMBER: US 09/924,340
PRIOR FILING DATE: 2001-08-06
PRIOR APPLICATION NUMBER: PCT/IB01/01715
PRIOR FILING DATE: 2001-08-06
PRIOR APPLICATION NUMBER: US 60/305,456
PRIOR FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: US 60/302,277
PRIOR FILING DATE: 2001-06-29
PRIOR APPLICATION NUMBER: US 60/298,698
PRIOR FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: US 60/293,574
PRIOR FILING DATE: 2001-05-25
NUMBER OF SEQ ID NOS: 112
SOFTWARE: JPatent
SEQ ID NO 58
LENGTH: 163
TYPE: PRT
ORGANISM: Homo sapiens
US-10-838-854-58

Query Match 100.0%; Score 163; DB 5; Length 163;
Best Local Similarity 100.0%; Pred. No. 1.2e-136;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGPPGKGTGHTGHLGPPKDCGKPPPGSTGRPGAEGEPGAMGPGCRPGPHVGP 60
Db 1 MGPPGKGTGHTGHLGPPKDCGKPPPGSTGRPGAEGEPGAMGPGCRPGPHVGP 60
Qy 61 PCQPGAGISAVGLKGDRTGATGSRGLAGLPGQGPFGPGPGPGYGTGKATGTPWGQGGIPG 120
Db 61 PCQPGAGISAVGLKGDRTGATGSRGLAGLPGQGPFGPGPGPGYGTGKATGTPWGQGGIPG 120
Qy 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQOYPPMKTMGPF 163
Db 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQOYPPMKTMGPF 163

Thu Dec 1 10:27:26 2005

us-09-924-340-58.rapbm

Page 4

Search completed: December 1, 2005, 09:02:32
Job time : 165 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 1, 2005, 08:49:06 ; Search time 11 Seconds
(without alignments)
70.954 Million cell updates/sec

Title: US-09-924-340-58
Perfect score: 163
Sequence: 1 MGPFGFKGTGHPGLPGRKG.....GAMPMEQGYPPMKTMKGRPG 163

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 26661 seqs, 4788334 residues

Word size : 100

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : Published Applications_AA_New:*

- 1: /cgn2_6/ptodata/2/pubppaa/US09_NEW_PUB.pep:*
- 2: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB.pep:*
- 5: /cgn2_6/ptodata/2/pubppaa/PCT_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/2/pubppaa/US10_NEW_PUB.pep:*
- 7: /cgn2_6/ptodata/2/pubppaa/US11_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/2/pubppaa/US60_NEW_PUB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description

No matches found

Search completed: December 1, 2005, 09:02:49
Job time : 11 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using bw model

Run on: December 1, 2005, 08:38:06 ; Search time 37 Seconds
(without alignments)
423.874 Million cell updates/sec

Title: US-09-924-340-58

Perfect score: 163
Sequence: 1 MGPPGFKGTGHPGLPGPKG.....GAMPMEQYPPMKTKMGPPG 163

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283416 seqs, 96216763 residues

Word size : 100

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database : PIR 80:*
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	ID	Description

No matches found

Search completed: December 1, 2005, 08:48:58
Job time : 38 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 1, 2005, 08:28:14 / Search time 236 Seconds
(without alignments)
487.293 Million cell updates/sec

Title: US-09-924-340-58

Perfect score: 163
Sequence: 1 MGPPGFGKGTGHPGLPDPKG.....GAMPMEQGYPPMKTKMGPRG 163

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 2166443 seqs, 705528306 residues

Word size : 100

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : UniProt_05.80.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description

No matches found

Search completed: December 1, 2005, 08:45:00
Job time : 237 secs

THIS PAGE BLANK (USPTO)